



## INTERNATIONAL APPLICATION PUBLISHED UNDER THE PATENT COOPERATION TREATY (PCT)

|   |           |   |
|---|-----------|---|
| <p>(51) International Patent Classification <sup>6</sup> :<br/>C12N 15/10, 15/12, 15/24, 15/62, 15/70</p>   | <p>A1</p> | <p>(11) International Publication Number: <b>WO 97/32017</b><br/>(43) International Publication Date: 4 September 1997 (04.09.97)</p>   |
| <p>(21) International Application Number: PCT/EP97/00931<br/>(22) International Filing Date: 26 February 1997 (26.02.97)<br/>(30) Priority Data: 96102852.9 26 February 1996 (26.02.96) EP<br/>(34) Countries for which the regional or international application was filed: DE et al.<br/>(71) Applicant (for all designated States except US): MORPHOSYS GESELLSCHAFT FÜR PROTEINOPTIMIERUNG MBH [DE/DE]; Frankfurter Ring 193a, D-80807 München (DE).<br/>(72) Inventors; and<br/>(75) Inventors/Applicants (for US only): ILAG, Vic [PH/DE]; Kporrstrasse 85, D-80807 München (DE). GE, Liming [CN/DE]; Portiastrasse 12, D-81545 München (DE).<br/>(74) Agent: VOSSIUS &amp; PARTNER; P.O. Box 86 07 67, D-81634 München (DE).</p>   |           | <p>(81) Designated States: CA, JP, US, European patent (AT, BE, CH, DE, DK, ES, FI, FR, GB, GR, IE, IT, LU, MC, NL, PT, SE).<br/><br/><b>Published</b><br/>With international search report.<br/>Before the expiration of the time limit for amending the claims and to be republished in the event of the receipt of amendments.</p> |
| <p>(54) Title: NOVEL METHOD FOR THE IDENTIFICATION OF NUCLEIC ACID SEQUENCES ENCODING TWO OR MORE INTERACTING (POLY)PEPTIDES</p>  |           |   |
| <p>(57) Abstract</p> <p>The present invention relates to methods for identifying nucleic acid sequences which encode two or more specific interacting peptides or proteins. Furthermore, the present invention relates to kits which may be used for the identification of nucleic acid sequences in accordance with the method of the present invention.</p> <div data-bbox="893 1123 1396 1858"> <p>General description of the polyphage principle</p> <p>The diagram illustrates the general description of the polyphage principle. It shows two phage libraries, Library 1 and Library 2, each containing a gene (tag gene and library gene) and an origin of replication. Library 1 also includes an antibiotic resistance gene. The diagram shows the process of infection, where Library 1 infects a cell, followed by Library 2, resulting in a co-infected cell containing both libraries.</p> </div> |           |   |

**FOR THE PURPOSES OF INFORMATION ONLY**

Codes used to identify States party to the PCT on the front pages of pamphlets publishing international applications under the PCT.

|    |                          |    |                                       |    |                          |
|----|--------------------------|----|---------------------------------------|----|--------------------------|
| AM | Armenia                  | GB | United Kingdom                        | MW | Malawi                   |
| AT | Austria                  | GE | Georgia                               | MX | Mexico                   |
| AU | Australia                | GN | Guinea                                | NE | Niger                    |
| BB | Barbados                 | GR | Greece                                | NL | Netherlands              |
| BE | Belgium                  | HU | Hungary                               | NO | Norway                   |
| BF | Burkina Faso             | IE | Ireland                               | NZ | New Zealand              |
| BG | Bulgaria                 | IT | Italy                                 | PL | Poland                   |
| BJ | Benin                    | JP | Japan                                 | PT | Portugal                 |
| BR | Brazil                   | KE | Kenya                                 | RO | Romania                  |
| BY | Belarus                  | KG | Kyrgyzstan                            | RU | Russian Federation       |
| CA | Canada                   | KP | Democratic People's Republic of Korea | SD | Sudan                    |
| CF | Central African Republic | KR | Republic of Korea                     | SE | Sweden                   |
| CG | Congo                    | KZ | Kazakhstan                            | SG | Singapore                |
| CH | Switzerland              | LI | Liechtenstein                         | SI | Slovenia                 |
| CI | Côte d'Ivoire            | LK | Sri Lanka                             | SK | Slovakia                 |
| CM | Cameroon                 | LR | Liberia                               | SN | Senegal                  |
| CN | China                    | LT | Lithuania                             | SZ | Swaziland                |
| CS | Czechoslovakia           | LU | Luxembourg                            | TD | Chad                     |
| CZ | Czech Republic           | LV | Latvia                                | TG | Togo                     |
| DE | Germany                  | MC | Monaco                                | TJ | Tajikistan               |
| DK | Denmark                  | MD | Republic of Moldova                   | TT | Trinidad and Tobago      |
| EE | Estonia                  | MG | Madagascar                            | UA | Ukraine                  |
| ES | Spain                    | ML | Mali                                  | UG | Uganda                   |
| FI | Finland                  | MN | Mongolia                              | US | United States of America |
| FR | France                   | MR | Mauritania                            | UZ | Uzbekistan               |
| GA | Gabon                    |    |                                       | VN | Viet Nam                 |

## **NOVEL METHOD FOR THE IDENTIFICATION OF NUCLEIC ACID SEQUENCES ENCODING TWO OR MORE INTERACTING (POLY)PEPTIDES**

The present invention relates to methods for identifying nucleic acid sequences which encode two or more specific interacting peptides or proteins. Furthermore, the present invention relates to kits which may be used for the identification of nucleic acid sequences in accordance with the method of the present invention.

Protein-protein interactions play an important role in all biological processes, from the replication and expression of genes to the morphogenesis of organisms (Lewin, B. 1994, *Genes V*. Oxford University Press). Methods for detecting protein-protein interactions have proved useful in understanding the basic mechanisms of different biological processes and the development of therapeutics. Detection of protein-protein interactions can be divided into two main categories: (i) physico-chemical based and (ii) genetic approaches (Phizicky, E.,M. & Fields, S. *Microbiological Reviews* 59 (1995) 94-123). Detection of protein-protein interactions by physico-chemical methods usually requires significant amounts of material, and more importantly, the identity of the proteins to be studied must be known. Recent developments in methods of mass spectrometry circumvent this problem but such suffer the disadvantage of requiring sophisticated equipment and expertise (Wang, R. & Chait, B.T., *Current Opinion in Biotech.* 5 (1994) 77-84). In contrast, genetic approaches provide an easy and powerful method of identifying protein-protein interactions without the need for pure material and specialized equipment, with the added advantage of higher throughput.

Different genetic approaches have been used to identify protein-protein interactions. The current method of choice is the yeast 2-hybrid system (Fields, S. & Song, O.K.,

Nature (London) 340, (1989) 245-246) which allows the identification of novel proteins that interact with a known protein.

Another popular genetic approach is the phage display system (Patent Application WO90/02809) whereby proteins are fused to a component of a surface protein of filamentous phage to allow selection for binding to a ligand of interest. The gene encoding the protein displayed on the surface of the phage is packaged inside the phage allowing the coupling of genetic information with the gene product. This allows the screening of "libraries" of proteins whereby the identity of the screened protein is deduced from the nucleic acid sequence of the phage. This technique has been extended by Winter et al. (Patent Application WO 92/20791) to produce libraries of multimeric members of a specific binding pair (e.g. combinations of VH and VL chains of an antibody) and select for functional specific binding pair members that can bind to the complementary specific binding pair member (e.g. antigen). Said libraries are constructed by combining two sub-libraries each encoding a collection of corresponding sub-units of said multimeric members (e.g. a library of VH chains is combined with a library of VL chains) wherein in principle each sub-unit out of the first sub-library is able to bind to each sub-unit out of the second sub-library non-specifically. Although this method has led to the identification of unique antibodies against particular antigens, it fails to provide a method for identifying two partners of a specific binding pair when both are unknown.

A unique version of phage display which relies on non-infective phage has recently been proposed (Duenas, M. & Borrebaeck, C. A. K., Bio/Technology 12 (1994) 999-1002; EP 0 614 989). A version of this system that led to the identification of proteins from a cDNA library that interacts with the jun protein has been described (Gramatikoff et al., Nucleic. Acids Res. 22 (1994) 5761-5762). The same principle has been also shown to work with an antibody-antigen system (Krebber et al., FEBS Letters 377 (1995) 227-231).



In spite of the power of all the aforementioned genetic selection approaches, they are limited to the selection of interacting binding entities from only a single genetically-diverse population (library vs. individual).

It would, however, be highly desirable to simultaneously identify binding entities and their specific binding partners in a library vs. library setting, wherein preferably at least two genetically diverse populations are involved. A solution to this technical problem, i.e. the identification of interacting entities and the respective nucleic acid sequences from more than one genetically diverse population (library vs. library) is neither provided nor suggested by the prior art. The present invention solves the above technical problem by providing the embodiments characterized in the claims. By using these embodiments, it has become possible to increase exponentially the rate at which (poly)peptide-(poly)peptide interactions are detected. The present invention may find applications in the field of functional genomics, whereby different proteins of unknown functions can be related with other proteins.

Accordingly, the present invention relates to a method for identifying a plurality of nucleic acid sequences, said nucleic acid sequences each encoding a (poly)peptide capable of interacting with at least one further (poly)peptide encoded by a different member of said plurality of nucleic acid sequences, comprising the steps of:

- (a) providing a first library of recombinant vector molecules containing genetically diverse nucleic acid sequences comprising a variety of nucleic acid sequences encoding (poly)peptides;
- (b) providing a second library of recombinant vector molecules containing genetically diverse nucleic acid sequences comprising a variety of nucleic acid sequences encoding (poly)peptides capable of interacting with further (poly)peptides as mentioned in step (a), wherein the vector molecules

4-

employed for the production of said recombinant vector molecules and/or the recombinant inserts display properties that are phenotypically distinguishable from those of the vector molecules and/or the recombinant inserts used in step (a) and wherein at least one of said properties displayed by each of said vector molecules and/or the recombinant inserts used in steps (a) and (b), upon the interaction of a (poly)peptide from said first library with a (poly)peptide from said second library together generate a screenable or selectable property;

- (c) optionally, providing additional libraries of recombinant vector molecules containing genetically diverse nucleic acid sequences comprising a variety of nucleic acid sequences encoding (poly)peptides capable of interacting with or causing interaction of (a) further (poly)peptide(s) as mentioned in step (a) and/or step (b), wherein the vector molecules employed for the production of said recombinant vector molecules and/or the recombinant inserts display properties that are phenotypically distinguishable from those of the vector molecules and/or the recombinant inserts used in steps (a) and (b) and, optionally, at least one of said properties displayed by said vector molecule and/or the recombinant inserts used in step (c) together with at least one of said properties displayed by either said vector molecule and/or said recombinant insert used in steps (a) and/or (b), upon the interaction of a (poly)peptide from said additional library with either a (poly)peptide from said first library and/or a (poly)peptide from said second library generate a screenable or selectable property;
- (d) expressing members of said libraries of recombinant vectors or nucleic acid sequences mentioned in steps (a), (b) and optionally (c), in appropriate host cells so that at least one interaction is established;
- (e) selecting for the generation of said screenable or selectable property representing the interaction of said (poly)peptides;

- (f) optionally, carrying out further selection, screening and/or purification steps; and
- (g) identifying said nucleic acid sequences encoding said (poly)peptides.

Thus, in the context of the present invention, the term "properties that are phenotypically distinguishable" relates alternatively to properties that are encoded by the vector molecule or to properties that are encoded by the recombinant insert or to both types of properties. As regards the vector-encoded properties, these may e.g. be resistance markers or requirements for special nutrients. It should be noted that the recombinant insert may comprise a nucleic acid portion encoding said property in addition to the nucleic acid portion responsible for the interaction.

In the context of the present invention, the term "different member " denotes a different entity which may be, but is not necessarily, structurally different.

Further, in the context of the present invention, the term "plurality" bears the meaning of "at least two".

The novel properties generated by the at least two recombinant inserts reflect the inventive principle of the present invention. That is, only if two (or more) (poly)peptides interact, for example, in a homo-dimeric or hetero-dimeric fashion, a screenable or selectable property is generated. The interaction between the two or more molecules may be a direct one or may be mediated indirectly. Examples for a direct interaction are the binding of an antibody encoded by a nucleic acid sequence from library 1 to a cDNA protein from library 2, the binding of a protein encoded by a nucleic acid sequence from cDNA library 1 to a protein from a cDNA library 2, as well as of an anti-idiotypic antibody encoded by a nucleic acid sequence from one of the libraries to a corresponding antibody encoded by a nucleic acid sequence from the other library. The nucleic acid sequences are preferably DNA and most preferably genes or parts thereof.

An example of an indirect interaction is the bridging of two (poly)peptides encoded by the two libraries which is mediated by a phosphorylating enzyme. Once the phosphorylation of one (poly)peptide encoded e.g. by library 1 is effected by the respective kinase, then this protein is capable of interacting with the second (poly)peptide encoded by library 2. The phosphorylating enzyme exemplifying this type of interaction may be encoded by a nucleic acid from (one of) the additional libraries and/or may be encoded by the genome of the host cell. Typically, the interaction of the two (poly)peptides forms a "bridge" of molecules, said "bridge" being detectable using an appropriate detection process. Conveniently, said bridge is detectable by a tag molecule that is associated with, encoded by or attached to one of the (poly)peptides encoded by library 1 or preferably 2.

Furthermore, the present invention relates to a method for identifying a plurality of nucleic acid sequences, said nucleic acid sequences each encoding a (poly)peptide capable of interacting with at least one further (poly)peptide encoded by a different member of said plurality of nucleic acid sequences, comprising the steps of:

- (a) expressing in appropriate host cells
  - (aa) nucleic acid sequences contained in a first library of recombinant vector molecules containing genetically diverse nucleic acid sequences comprising a variety of nucleic acid sequences encoding (poly)peptides;
  - (ab) nucleic acid sequences contained in a second library of recombinant vector molecules containing genetically diverse nucleic acid sequences comprising a variety of nucleic acid sequences encoding (poly)peptides capable of interacting with further (poly)peptides as

mentioned in step (aa), wherein the vector molecules employed for the production of said recombinant vector molecules and/or the recombinant inserts display properties that are phenotypically distinguishable from those of the vector molecules and/or the recombinant inserts used in step (aa) and wherein at least one of said properties displayed by each of said vector molecules and/or the recombinant inserts used in steps (aa) and (ab), upon the interaction of a (poly)peptide from said first library with a (poly)peptide from said second library together generate a screenable or selectable property;

- (ac) optionally, nucleic acid sequences contained in additional libraries of recombinant vector molecules containing genetically diverse nucleic acid sequences comprising a variety of nucleic acid sequences encoding (poly)peptides capable of interacting with or causing interaction of (a) further (poly)peptide(s) as mentioned in step (aa) and/or step (ab), wherein the vector molecules employed for the production of said recombinant vector molecules and/or the recombinant inserts display properties that are phenotypically distinguishable from those of the vector molecules and/or the recombinant inserts used in steps (aa) and (ab) and, optionally, at least one of said properties displayed by said vector molecule and/or the recombinant inserts used in step (ac) together with at least one of said properties displayed by either said vector molecule and/or said recombinant inserts used in steps (aa) and/or (ab), upon the interaction of a (poly)peptide from said additional library with either a (poly)peptide from said first library and/or a (poly)peptide from said second library generate a screenable or selectable property;

so that at least one interaction is established;

- (b) selecting for the generation of said screenable or selectable property representing the interaction of said (poly)peptides;
- (c) optionally, carrying out further screening, selection and/or purification steps; and
- (d) identifying said nucleic acid sequences encoding said (poly)peptides.

In a preferred embodiment of the method of the present invention, said screenable or selectable property is expressed extracellularly.

This embodiment is conveniently employed in a number of laboratories which would make use of rather conventional methodology of the extracellular detection of such properties, e.g. by column chromatography wherein the e.g. screenable tag is retained, in combination with e.g. plaque purification techniques, which allow the further purification of the cells that were originally enriched by e.g. the column chromatography step.

In a further preferred embodiment of the method of the present invention, said recombinant vector molecule in step (a)/(aa) (the step identified after the slash refers to the corresponding step of the second embodiment of the method of the invention identified hereinabove) gives rise to a replicable genetic package (RGP) displaying said (poly)peptides at its surface. In this context, the term replicable genetic package (RGP) refers to an entity, such as a virus or bacteriophage, which can be replicated following infection of a suitable host cell. In the case of bacteriophage, for example, the collection of nucleic acid sequences can be inserted into either a phage or phagemid vector in frame with a component of the phage coat, such as gene III, resulting in display of the encoded binding entities on the surface of the phage. Particularly preferred as a

recombinant vector molecule is a recombinant phage, phagemid or virus, wherein said phage is most preferably

- (a) one of the class I phage fd, M13, If, Ike, ZJ/2, Ff;
- (b) one of the class II phage Xf, Pf1, and Pf3;
- (c) one of the lambdoid phages, lamda, 434, P1;
- (d) one of the class of enveloped phages, PRD1; or
- (e) one of the class paramyxo-viruses, orthomyxo-viruses, baculo-viruses, retro-viruses, reo-viruses and alpha-viruses.

In a further preferred embodiment of the method according to the invention, said selection step (e)/(b) is carried out by selecting polyphage comprising the interacting (poly)peptides. Polyphage contain more than one copy of phage genomic DNA. They occur naturally at a low to moderate frequency when a newly forming phage coat encapsulates two or more single-stranded DNA molecules. In the case of the present invention, the polyphage which are formed will contain at least two phage genomes, which may either (i) both be representatives of library 1, or (ii) both be representatives of library 2, or (iii) be representatives of each of library 1 and library 2, or (iv) be a combination of (i) to (iii) with at least one member of one of the additional libraries. The efficiency of polyphage production can be increased by the introduction of appropriate mutations into the phage genome, as is well known to those skilled in the art (see, for example, Lopez, J. and Webster, R.E., *Virology* 127 (1983), 177-193, Bauer, M. and Smith, G.P., *Virology* 167 (1988) 166-175, or Gailus, V. et al., *Res. Microbiol.* 145 (1994) 699-709).

In a further preferred embodiment of the method of the invention, said screenable or selectable property is connected to the infectivity of said RGP.

In this embodiment, use is made of the possibility that the infectivity of e.g. a bacteriophage can be manipulated, said infectivity being directly correlated with the interaction of said (poly)peptides.

In a most preferred embodiment of the method of the present invention, said RGP is encoded by said recombinant vector used in step (a)/(aa) and rendered non-infective and infectivity of said RGP is restored by interaction of said (poly)peptide of step (a)/(aa) with the (poly)peptide of step (b)/(ab) and/or (c)/(ac), said (poly)peptide of step (b)/(ab) and/or (c)/(ac) being fused to a domain that confers infectivity to said RGP.

In a further most preferred embodiment of the method of the invention, said RGP is rendered non-infective by modification of a genetic sequence which encodes a surface protein necessary for the RGP's binding to and infection of a host cell.

These preferred and most preferred embodiments of the method of the present invention relating to the infectivity of the RGP serve as an alternative to the use of the screenable tag. In these embodiments, advantage can be taken of the phenomenon of selective infection (Krebber et al., FEBS Letters 377 (1995) 227-239). While the screenable tag enables physical separation of molecules from others in the population, the use of selective infection enables positive selection for the interacting pair. This phenomenon relies on the use of a construct which can selectively restore infectivity to phage which have been rendered non-infective by, for example, deletion of all but the C-terminus of the gene III protein. Use of such phage for displaying library 1 gives non-infectious phage carrying the binding entity. Co-expression with library 2 allows interactions between binding entities and binding partners to be established, as described above. Although the phage which carry the binding entity-binding partner pair are non-infective, infectivity can be restored if, in place of the screenable tag referred to above, an infectivity protein is used. In this context, the term infectivity protein refers to a substance which, when associated with the phage, can enable it to penetrate a bacterial host, where it is subsequently replicated. An example of an infectivity protein is the N-terminus (at least the first 220 amino acids) of gene III protein of the filamentous bacteriophage.



The infectivity protein confers on those phage which carry it, the ability to be replicated. Thus, only those phage which carry the binding entity/partner pair are replicated. Purification of hybrid phage containing genes from both libraries 1 and 2 then relies e.g. on the use of two selectable markers as indicated above. The genes in the phage can then be identified using methodology well known to those skilled in the art.

An additional preferred embodiment of the present invention relates to a method, wherein said recombinant vector molecules in step (a)/(aa) give rise to a fusion protein which is expressed on the surface of a cell, preferably a bacterium.

These fusion proteins, upon interaction with a suitable binding partner from library 2 connected e.g. with a screenable tag can be detected on the surface of host cells which may be, for example, bacteria, yeast, insect cells or mammalian cells. The display of fusion proteins on bacterial surfaces per se is well known in the art. Thus, lipoproteins (Lpp), outer membrane proteins A (OmpA), and flagella have been used to target antibodies and peptides to the cell surface of E.coli. Fuchs et al., Bio/Technology 9 (1991) 1369-1372, WO93/01287, presented a single chain antibody on the surface of E.coli as a fusion protein with the N-terminus of the peptidoglycan-associated lipoprotein. The antibody was visualized by the binding of fluorescently labeled antigen and fluorescently labeled antibodies directed to the linker peptide of the displayed single chain antibody. Francisco et al., Proc. Natl. Acad. Sci. USA 90 (1993) 10444-10448, and Georgiu, G. et al., WO93/10214, displayed antibodies on the E.coli surface by fusing the N-terminus of a single chain antibody to the C-terminus of OmpA while the N-terminus of OmpA was fused to the signal sequence and the first nine amino acids of Lpp. Binding of a fluorescently labeled antigen to the OmpA-antibody fusion protein was detected by FACS. Klauser (WO 95/17509) transferred the IgA protease system from Neisseria to E.coli to facilitate display of antibodies. Integration of the beta-domain of the IgA protease precursor into the outer membrane lead to the transport of the

protease domain across the membrane followed by autoproteolytic release into the medium. Antibodies linked to the beta-domain of IgA protease are therefore presented on the surface of bacteria. Further, Lu, Z. et al., Bio/Technology 13 (1994) 366-371, described a system for displaying peptides on the surface of the bacterium by fusing it to thioredoxin and the bacterial flagella, to screen for peptide mimics of the epitope for an anti-IL-8 antibody.

The further identification of the desired nucleic acid molecule encoding the interacting (poly)peptides may then be effected by methods known in the art, e.g. by purifying host cells displaying a tag on their surface and further by antibioticum-based selection techniques, DNA purification and sequencing.

In a particularly preferred embodiment of the method of the present invention, said bacterium is *Neisseria gonorrhoe* or *E.coli* and said fusion protein consists of at least a part of a flagellum, lam B, peptidoglycan-associated lipoprotein or the Omp A protein and said (poly)peptide.

As has been repeatedly pointed out hereinabove, a tag connected to the (poly)peptide encoded by library 2 can conveniently be used in the identification strategy of the desired nucleic acid sequences. Accordingly, in a further preferred embodiment of the method of the invention, said (poly)peptides encoded by said recombinant vector molecules of step (b)/(ab) or (c)/(ac) are linked to at least one screenable or selectable tag. In this context, the term screenable or selectable tag refers to a short sequence of amino acids which can be recognized and bound by a particular substance. Tags are commonly used for the purification of biomolecules: examples are His(n), where n = 4-6 which can be bound either by Ni, or a specific antibody, and the flag and myc tags which are recognized by appropriate antibodies. In either of these cases, the tag can be encoded as a C-terminal fusion to all binding partners in library 2. In accordance with the present invention, the tag can be used to isolate e.g. the polyphage referred to

above. Thus, the interaction between the phage-bound binding entity, and its interacting binding partner, establishes a connection between the phage particle and the screenable or selectable tag. This feature can be exploited in a step which relies on e.g. affinity chromatography to isolate the polyphage carrying the interacting molecules. In a final step, those polyphage which carry two distinct nucleic acid molecules and preferably genes (encoding binding entity and binding partner) can be separated from those carrying only one of the two genes e.g. by selection based on transduction or different selectable markers (e.g. antibiotic resistance) present in the individual genomes. In this way, the genes which encode the two interacting molecules can be identified.

A most preferred embodiment of the present invention relates to a method wherein said screenable or selectable tag is encoded by said recombinant vector of step (b)/(ab) or (c)/(ac).

A further most preferred embodiment of the present invention relates to a method wherein said screenable or selectable tag is selected from the list His(n), myc, FLAG, malE, thioredoxin, GST, streptavidin, beta-galactosidase, alkaline phosphatase T7 gene 10, Strep-tag and calmodulin. These screenable tags are all well known in the art and are fully available to the person skilled in the art.

In an additional particularly preferred embodiment of the method of the invention, said screenable or selectable tag is encoded by the genome of the host cell.

An example for this embodiment is an anti-Fc-receptor specific antibody that is expressed by the host cell and could function as an additional bridge in e.g. purification by column chromatography. Another example of this embodiment is an enzyme produced by the host cell that creates a tag such as a phosphorylation on (poly)peptides of the second library without destroying the interaction of (poly)peptides of step (b)/(ab)

with (a)/(aa) so that the modification caused by the enzyme is now the screenable or selectable tag.

In a further preferred embodiment of the method of the invention, said (poly)peptides encoded by the nucleic acid sequences of said additional libraries of step (c)/(ac) cause the interaction of said (poly)peptides of steps (a)/(aa) and (b)/(ab) via phosphorylation, glycosylation, methylation, lipidation or farnesylation of at least one of said (poly)peptides of steps (a)/(aa) and (b)/(ab).

An additional preferred embodiment of the invention relates to a method wherein said host cells in step (d)/(a) are spatially addressable, and the nucleic acid sequences mentioned in step (g)/(d) are retrieved from the corresponding spatially addressable host cell.

In the context of the present invention, the term "spatially addressable" refers to a situation where the individual cells harboring one of the potential combinations of members of the first, second and optionally additional libraries are identifiable by their relative position, e.g. by their position on a master plate. The screening or selection may, for example, be performed either with single clones derived from the master plate, or on a replica plate, thus maintaining the connection between the screenable or selectable property and the information contained in the host cell on the master plate.

An additional preferred embodiment of the invention relates to a method wherein said screenable or selectable property is expressed intracellularly.

Particularly preferred is a method wherein said screenable property is the transactivation of the transcription of a reporter gene such as beta-galactosidase, alkaline phosphatase or nutritional markers such as his3 and leu or resistance genes

giving resistance to an antibiotic such as ampicillin, chloramphenicol, kanamycin, zeocin, neomycin, tetracycline, or streptomycin.

Furthermore, use can be made of the yeast 2-hybrid system referred to hereinabove or the interaction trap system (Brent et al., EP-A 0 672 131) or of a prokaryotic version analogous to the above recited systems, utilizing the toxR system of *Vibrio cholerae* (Fritz, H.-J. et al., EP-A 0 630 968). It is within the skills of the person skilled in the art to combine further screening systems known in the art with the method of the present invention.

In a further preferred method of the present invention, said recombinant vectors of step (a)/(aa), (b)/(ab) and (c)/(ac) comprise recombination promoting sites and in said step (e)/(b) recombination events are selected for, wherein said nucleic sequences encoding said (poly)peptides of step (a)/(aa), said nucleic acid sequences encoding said (poly)peptides of step (b)/(ab) and optionally said nucleic acid sequences encoding said (poly)peptides of step (c)/(ac) are contained in the same vector. In this approach, the two genes can be coupled in a single vector, and packaged in a phage of standard size, if appropriate recombination sites are incorporated in the vectors carrying libraries 1 and 2. Again, the phage which carry both nucleic acid sequences and genes are purified with the use of e.g. the screenable tag. If recombination is used to couple the genes from the two libraries, some of the hybrid progeny phage will contain nonrecombinant genomes, since site-specific recombination is not very efficient. However, the hybrid phage can be selected by re-infection of host cells that do not contain library 2 followed by another round of selection of the screenable tag.

In a particularly preferred embodiment of the method of the invention, said recombination events are mediated by the site-specific recombination mechanisms Cre-lox, attP-attB, Mu gin or yeast flp.

In a further particularly preferred embodiment of the method of the invention, said recombination promoting sites are restriction enzyme recognition sites and said recombination event is achieved by cutting the recombinant vector molecules mentioned in steps (a)/(aa), (b)/(ab) and optionally (c)/(ac) with at least two different restriction enzymes and effecting recombination of the nucleic acid sequences contained in said vectors by ligation.

The invention relates in an additional preferred embodiment to a method wherein said identification of said nucleic acid sequences is effected after the selection step (e)/(b) via PCR and preferably sequencing of said nucleic acid sequences after said PCR.

After said selection step (e)/(b), PCR can be carried out with the enriched desired product, conveniently using primers that hybridize to the vector portion of the recombinant vector molecule. Sequencing of the PCR-product may then be carried out according to conventional methods.

In a further preferred embodiment of the method according to the invention, said recombinant vectors of step (a)/(aa), (b)/(ab) and/or (c)/(ac) comprise at least one gene encoding a selection marker.

Said genes encoding said selection markers are preferably different in each of the vectors of step (a)/(aa), (b)/(ab) and/or (c)/(ac), i.e. said vectors comprise genes encoding different selection markers. Said selection markers can conveniently be used for the further purification envisaged in step (f)/(c). For example, a polyphage comprising two members of each library 1 and 2 can be selected for on the basis of a double resistance to antibiotics. Also, a successful recombination event may create a new recombinant vector carrying both nucleic acid molecules from library 1 and 2 as well as genes encoding different selection markers. Again, the selection for a twofold resistance will assist in the identification of the desired product.

In a particularly preferred embodiment of said method, said selection marker is a resistance to an antibiotic, preferably to ampicillin, chloramphenicol, kanamycin, zeocin, neomycin, tetracycline or streptomycin.

A further preferred embodiment of the present invention relates to a method wherein said host cells are F' and preferably E.coli XL-1 Blue, K91 or its derivatives, TG1, XL1kan or TOP10F.

In a particularly preferred embodiment of the present invention, said RGP's are produced with the use of helper phage taken from the list R408, M13k07 and VCSM13, M13de13, fCA55 and fKN16 or derivatives thereof.

Further preferred is a method wherein at least one of said genetically diverse nucleic acid sequences encode members of the immunoglobulin superfamily.

Said method is particularly preferred, if said genetically diverse nucleic acid sequences encode a repertoire of immunoglobulin heavy or light chains.

In an additional preferred embodiment of the present invention, in said method said genetically diverse nucleic acid sequences are generated by a mutagenesis method. Various mutagenesis methods are well known to the person skilled in the art and need not be described in here in any further detail.

The present invention relates in an additional preferred embodiment to a method in which said genetically diverse nucleic acid sequences are generated from a cDNA library.

In a final preferred embodiment of the method of the invention, said nucleic acid sequences are genes or parts thereof.

As used herein, the term "parts thereof" relates to parts of genes that encode a product that is capable of interacting with a product encoded by any of the other libraries. Thus, it is well known that various proteins are comprised of different domains. Only one of said domains may be capable of interacting with a different (poly)peptide. Such a domain might be encoded by a part of said gene in accordance with the present invention.

The invention also provides for identifying genes encoding more than two interacting peptides or proteins. This can be achieved by using additional vectors encoding genetically diverse additional nucleic acids by an extension of the method described above. As previously, the presence of either a screenable tag or an infectivity protein is used to purify phage carrying genes which encode the components of the complex. Again, the genes in the phage can then be sequenced using methodology well known to those skilled in the art.

Additionally, the present invention relates to a kit comprising at least

- (a) a recombinant vector molecule as described in step (a)/(aa) or a corresponding vector molecule;
- (b) a recombinant vector molecule as described in step (b)/(ab) or a corresponding vector molecule; and, optionally;
- (c) at least one further recombinant vector molecule as described in step (c)/(ac) or a corresponding vector molecule.

As a rule, if recombinant vector molecules are comprised in said kit, they will comprise a library of nucleic acid molecules. In other words, the kit of the invention will contain a plurality of different recombinant vector molecules.



## Legends to Figures and Tables

**Figure 1:** General description of the polyphage principle

- a) transform to *E. coli* hosts
- b) infect host containing library1 with helper-phage to package library1 into phage
- c) infect cells containing library2 with phages containing library1 leading to cells harboring members of library1 and library2; the presence of library1 and library2 is selected by the presence of the 2 antibiotic resistance markers
- d) expression of library1 and library2-tag gene products
- e) infect cells with engineered helper-phage to induce polyphage production

Note 1: Polyphage does not discriminate which genome to package therefore the possibilities resulting from step e) arise in an infected cell. To select for the polyphage containing the right packaged genomes the subsequent step is required

f) select for tag e.g., infectivity-mediating protein, in which case ability to infect is selected and

g) select for ability to confer resistance to 2 antibiotics to infected cells

Note 2: Only polyphages that satisfy f) + g) represent phages that display the correct interacting pair and the corresponding genetic information

**Figure 2:** Co-transformation of two phagemids, polyphage formation and selection via His-tag: general description

A, B: libraries of phagemids, preferably with different resistance markers; A: fusions to gIIIp; B: fusions to tag (His); after co-transformation phage production leading to a phage population displaying cognate pairs (left part of the Figure) or not (right part), after selection infection of host cells, selection for double-resistance

Alternative methods include the infection of cells harbouring a plasmid- or phagemid-based library B with a phage library A (prerequisite again: interference-resistant constructs).

**Figure 3:** pBS vector series: functional map and sequence of pBS13

**Figure 4:** Co-existence of phagemids: results of restriction digest

Restriction analysis of clones of double resistances (Amp/Cm). R1: pIG10.3, *Xba*/*Sca*I; R2: pBS13, *Xba*/*Sca*I, R1+R2: R1 and R2 are mixed in approx. equal proportion; M1: marker  $\lambda$ : *Bst*EII; M2: marker pBR322: *Msp*I; 1 to 10: randomly picked clones: *Xba*/*Sca*I

**Figure 5:** Phagemid vector pYING1-C1: functional map

containing the fos peptide. The corresponding vectors pYING1-C2 and pYING1-C3 contain instead of fos the p75 and the IL16 peptides, respectively

**Figure 6:** Phagemid vector pYANG3-A: functional map

containing the jun peptide. The corresponding vectors pYANG3-Ape2, pYANG3-Ape3, and pYANG3-Ape10 contain instead of jun the p75-binding peptides pe2, pe3, and pe10, respectively

**Figure 7:** Analysis of selected clones (see Table 2):

7.a: Restriction digest of clones before and after selection

R: pYANG3-Ape2: *Xba*I; M1: marker  $\lambda$ : *Bst*EII; M2: marker pBR322: *Msp*I;  $\alpha$ /1 to 10: randomly picked clones before selection: *Xba*I/*Hind*III;  $\beta$ /1 to 10: randomly picked clones after selection: *Xba*I/*Hind*III; size expected: jun-gIII: 745 bp; fos: 256 bp; p75: 577 bp; IL-16: 502 bp

7.b: PCR reaction of clones after selection with primers OPEP5L and OGIII3

R1: pYANG3-A as template; R2: pYANG3-Ape2 as template; M: marker  $\lambda$ : *Bst*EII;  $\beta$ /1 to 10: randomly picked clones after selection as templates

**Figure 8:** Phagemid vector pING1-C1: functional map

containing the His-tag peptide. The corresponding vector pING3-C1 contains an additional FLAG epitope; pING1-C2 and pING3-C2 contain

the Strep-tag instead of His-tag, with pING3-C2 containing an additional FLAG epitope.

- Figure 9:** Phagemid vector pONG3-A: functional map  
for the generation of phage-display libraries (gIII fusions)
- Figure 10:** Co-transformation of phage and plasmid, polyphage formation and selection via SIP: general description  
fA: library A in phage construct; B: library B, library members fused to IMP; preferably different resistance markers on phage and plasmid; after co-transformation production of phages; in the case of cognate-pair interaction formation of infectious phages; selection; by plating on double-resistance identification of polyphage particles.
- Figure 11:** Phage vector fhag1A: functional map  
for phage-display of the  $\alpha$ -HAG scFv
- Figure 11a:** CAT gene module: functional map and sequence
- Figure 12:** Phage vector fjun1A: functional map  
for phage-display of the jun peptide
- Figure 13:** Phage vector fjun1B: functional map  
for phage-display of the jun peptide
- Figure 14:** Phage vector fpep3\_1B: functional map  
for phage-display of the peptide pe3 binding to the intracellular domain of p75
- Figure 15:** Phage vector fNGF\_1B: functional map  
for phage-display of NGF
- Figure 16:** Plasmid pUC19/IMPhag: functional map  
containing fusion of HAG peptide to the N-terminal domains of gIIIp (IMP)
- Figure 17:** Plasmid pUC18/IMPp75: functional map  
containing fusion of the intracellular domain of p75 to the N-terminal domains of gIIIp (IMP); pUC18/IMPfos contains the fos peptide instead of the intracellular domain of p75

**Figure 18:** Plasmid pUC18/IMPIL16: functional map

containing fusion of IL16 to the N-terminal domains of gIIIp (IMP)

**Figure 19:** Analysis of selected clones (see Table 3)

Lane 1: marker  $\lambda$ : *Bst*Ell; lanes 2 to 20: polyphage transductant clones #1 to #19 digested with *Xba*/*Hind*III; f..\_1b: fragment of phage vector after digest; pUC18: fragment of plasmid after digest;  $\alpha$ -HAG: fragment containing anti-HAG scFv fused to gIIIc; IMP-p75 and IMP-HAG: fragment containing IMP fused to p75, and IMP-HAG peptide, respectively; pep3-gIIIc: fragment containing pep3 fused to gIIIc (s: short version)

**Figure 20:** Co-transformation of phagemids, *in vivo* recombination and selection via His-tag: general description

A, B: libraries of phagemids; preferably with different resistance markers; A: fusions to gIIIp; B: fusions to tag (His); both constructs containing recombination-promoting sites (\*) such as *lox*/*loxP*; after co-transformation and recombination production of phages; selection via Ni-NTA; re-infection of host cells, selection for double-resistance

**Figure 21:** *In vitro* recombination and selection via His-tag: general description

A, B: libraries of phagemids; preferably with different resistance markers; A: fusions to gIIIp; B: fusions to tag (His); both constructs containing corresponding recognition sites for restriction enzymes (+/o); after digest and co-ligation transformation and production of phages; selection via Ni-NTA; re-infection of host cells, selection for double-resistance

**Figure 22:** Phage vector fjunhag: functional map for phage display of the jun peptide

**Figure 23:** Spatial *in vivo* SIP: general description

After transformation or co-transformation according to any of the methods described above, a master plate is made. From that phages secreted from individual clones can be analyzed individually (top), or a replica (migration of secreted phages through filter disc) can be made whereon selection for the presence of a tag or infectivity can be performed. By going back to the

master-plate, the information for selected cognate interacting pairs can be retrieved without requiring recombination and/or polyphage production.

**Figure 24:** *E. coli* display: general description

A, B: libraries of phagemids; preferably with different resistance markers;  
A: fusions to *E. coli* surface-display protein; B: fusions to tag (His); after co-transformation expression of constructs; surface-display; in the case of cognate interaction taking place, display of tag on the surface of the host cell; selection

**Figure 25:** pTERMsc2H10myc3sCAM: functional map and sequence

**Table 1:** Phagemids constructed for Experiments 2 and 3

**Table 2:** Results of Experiment 2 (see Figure 7)

- 2.a: Combination of phagemids present in initial library ( $\alpha$ )
- 2.b: Combination of phagemids present after selection ( $\beta$ )

**Table 3:** Results of Experiment 4 (see Figure 19)

- 3.a: Identification of phage/plasmid present in individual clones
- 3.b: Test for infectivity of individual clones

The examples illustrate the invention.

**Example 1: General description of the polyphage principle (Figure 1)**

The binding entities which comprise library 1 may be peptides or proteins, and are encoded by a genetically diverse collection of first nucleic acid sequences. These nucleic acid sequences are inserted into a first vector which allows for display of the encoded binding entities on the surface of a replicable genetic package. For the purposes of subsequent selection, the first vector should also carry a gene encoding a selectable marker, such as an antibiotic resistance. The binding partners which comprise library 2 may be peptides or proteins, and are encoded by a genetically diverse collection of second nucleic acid sequences which are inserted into a second vector. By way of example, this second vector may be a plasmid, or even a phage or phagemid, in which case the origin of replication should be distinct from that of the first vector. For the purposes of subsequent selection, the second vector should also carry a gene encoding a selectable marker, such as an antibiotic resistance, preferably distinct from that present in the first vector. To facilitate purification of the complex to be formed between any binding entity-binding partner pair, a screenable tag can be conveniently attached to members of library 2.

The two genetically diverse collections of nucleic acids are then introduced into a population of host cells in such a way that encoded libraries 1 and 2 can be expressed. This can be achieved by either (i) co-transformation of the two vectors, or, as actually shown in the figure, (ii) packaging one of the collections of nucleic acids into a vector (such as a bacteriophage) which can be used to infect with high efficiency a population of cells into which the complementary collection of nucleic acid has been introduced. The result is a population of cells in which individual cells carry representatives of each library.

Expression of the two collections of nucleic acids results in the production of pairs of molecules, one from each library, in the host cells. In each case, one or more members

of the library of binding entities is incorporated into the coat of an RGP. In some cells, an interaction will be established between a binding partner on the surface of the RGP and a binding partner expressed from library 2. When such an interaction is established, the RGP therefore carries both the binding entity and the binding partner.

The RGPs displaying such an interaction can then be further purified with the help of polyphage and differing selection markers, as has been discussed hereinabove. After such selection, the nucleic acid sequences encoding one or both binding partners can be conveniently identified by methodology known in the art, such as DNA sequencing.

### **Example 2: Co-transformation of phagemids with same *E. coli* origin of replication, polyphage formation, and selection of correct pairing interactions via His-tag**

#### **2.1: Principle (see Figure 2)**

To demonstrate that polyphage formation allows the retrieval of the genetic information for cognate protein pairs selected using a tag fused to one member of the protein pair, two separate, small libraries in phagemid vectors are constructed.

**2.2: Test of co-existence of phagemids with the same *E. coli* origin of replication:**  
Prerequisite for the formation of polyphage particles containing two different phagemids is that the different phagemid vectors can co-exist in the host cell.

The vector pBS13 is a derivative of the vector (Krebber *et al.*, 1996) containing a chloramphenicol-resistance gene instead of the kanamycin-resistance gene and a beta-lactamase gene cassette instead of the 2H10-gIII fusion gene, and can be assembled by standard methods starting from pto2H10a3s. Figure 3 contains the functional map and the sequence of pBS13. pIGHAG1A (see Example 4.2.1.f) is digested with *Xba*I and *Hind*III. The 1.3 kb fragment containing the anti-HAG gene fused with the C-

terminal domain of filamentous phage pIII protein is isolated and ligated with a pre-digested phagemid vectors pIG10.3, and pBS13 (XbaI-HindIII) to create the vectors pIG10.3-scFv(anti-HAG) (Ap<sup>R</sup>) and pBS13-scFv(anti-HAG) (Cm<sup>R</sup>), respectively. The vectors are used to transform competent XL-1 Blue cells and selected on LB plates containing Amp/Cm/Tet and glucose (20 mM).

The phagemids from clones of double-resistant colonies (Amp/Cm) are isolated. The restriction digestions indicate the co-isolation of both phagemids from the single colonies (Figure 4).

### 2.3: Design of libraries A and B:

**Library A** contains three cyclic peptides each binding to the intracellular domain of the low affinity nerve growth factor (NGF) receptor (see Example 4), and a leucine zipper domain derived from the jun transcription factor, all N-terminally fused to the C-terminal domain of gIII from filamentous phage.

**Library B** encodes 3 members, namely the leucine zipper domain of the fos transcription factor which heterodimerizes with jun via this domain, the intracellular domain of the NGF receptor p75, and, as a negative control which does not interact with library A members, IL-16, all fused at the N-terminus with a His<sub>6</sub>-peptide as tag (Hochuli *et al.*, 1988; Lindner *et al.*, 1992).

The cognate pairings are from the interaction between jun and fos (Cramer and Suter, 1993), and p75 and selected cyclic peptides (see Example 4). A non-cognate pairing would occur among the non-cognate pairs mentioned and among jun, or one of the cyclic peptides, and IL-16.

### 2.4: PCR amplification of the individual constructs

Fos, N-terminus fused to His<sub>6</sub>, is PCR amplified using pOK1 (Gramatikoff *et al.*, 1994) as template and oligonucleotides OFOS-5 and OFOS-3 as primers, where His<sub>6</sub> is



encoded in the OFOS-5 primer. Jun is PCR amplified using pOK1 as template and oligonucleotides OJUN-5 and OJUN-3 as primers.

OFOS-5 5'- GGGGATATCCACCACCACCACCACCTGCGGTGGTCTGACC

OFOS-3 5'- GGGGAATTCCAACCAACCGTGTGCCG

OJUN-5 5'- GGGGATATCGGTGGTCGGATCGCC

OJUN-3 5'- GGGGAATTCACCACCGTGGTTCATGAC

The hot-start procedure is used. A step-wise touch-down PCR is applied: 92°C, 1 min; 58-52°C,  $\Delta T = 2^\circ\text{C}$ , 1 min; 72°C, 1 min. This is followed by 26 cycles (92°C, 1 min; 52°C, 1 min; 72°C, 1 min).

The PCR products are purified using QIAquick kit (Qiagen) and eluted in ddH<sub>2</sub>O. They are then overnight digested with *EcoRI* and *EcoRV*.

The p75 fragment is also PCR amplified using pUC18-IMPp75 (see Example 4) as template and oligonucleotides OP75-5 (where His<sub>6</sub> is encoded) and OP75-3 as primers:

OP75-5 5'- GGGGATATCCACCACCACCACCACCAAGAGGTGGAACAGC

OP75-3 5'- GGGGAATTCCTACTGGGGATGTGGCAG

The same PCR and restriction digestion conditions as above are applied.

The IL-16 fragment is amplified from the cDNA clone pcDNA3-ILHu1 (M. Baier, Paul Ehrlich Institute, Germany; Baier *et al.*, 1995; Bannert *et al.*, 1996), using OIL16-5 (where His<sub>6</sub> is encoded) and OIL16-3 as primers.

OIL16-5 5'- GGGGATATCCACCACCACCACCACCCCGACCTCAACTCCTC

OIL16-3 5'- GGGGAATTCGGAGTCTCCAGCAGCTG

The same PCR and restriction digestion conditions as above are applied.

In all cases, the fragments are readily amplified and digested.

## 2.5: Cloning Into Intermediate vectors

The digested PCR fragments are gel-purified (QIAquick kit, Qiagen) and eluted into TE buffer. The *EcoRV/EcoRI* fragment of pIG1 vector (Ge *et al.*, 1995) is also isolated. The digested PCR fragments of fos, p75, and IL-16 are ligated into the vector fragment, and the ligated vectors transformed into TG1 cells.

The constructs in the pIG1 vector contains the OmpA signal sequence fused in-frame with the constructs.

The correct clones are screened and confirmed by sequencing. They are then *XbaI/HindIII* digested, and the fragments are isolated.

## 2.6: Cloning into the expression vectors

The isolated fragments from 2.3 are inserted into pBS13 also excised with *XbaI/HindIII*, resulting in vectors **pYING1-C1** (Fos), **pYING1-C2** (p75), **pYING1-C3** (IL-16) (see Figure 5). The fragment containing jun is cloned into pIG10.3 vector *via EcoRV/EcoRI* resulting in **pYANG3-A** (see Figure 6). The anti-p75 peptides pe2, pe3 and pe10 (see Example 4) are cloned into pIG10.3 *via XbaI/HindIII*, resulting in vectors **pYANG3-Ape2**, **-Ape3** and **-Ape10**, respectively (see Figure 6).

## 2.7: Selection of correct pairing *via* His-tag

TG1 cells are transformed with the combination of pYANG3-A + pYING1-C1, or pYANG3-A + pYING1-C2, or pYANG3-A + pYING1-C3, or (pYANG3-Ape2, -Ape3 and -Ape10) + pYING1-C1, or (pYANG3-Ape2, -Ape3 and -Ape10) + pYING1-C2, or (pYANG3-Ape2, -Ape3 and -Ape10) + pYING1-C3, thus creating all possible combinations separately to ensure the presence of each of them in the selection experiment. The transformed cells are plated on ampicillin/chloramphenicol-containing LB agar plates, and colonies with double resistance ( $Ap^R/Cm^R$ ) are selected.

The colonies are scraped off the plates and used to inoculate 2xYT medium (Amp/Cm) and shaken at 37°C for 3 hrs. The cultures are induced (1 mM IPTG) at 30°C for 1 hr and infected with R408 (Stratagene) at 37°C for 30 min. The cultures are shaken at RT for 3 hrs, kanamycin is added and shaking continued at RT overnight.

The phage particles are harvested from the overnight cultures, mixed and PEG-precipitated. The phages are directly selected on immobilized Ni-NTA (NI-NTA HisSorb Strips, Qiagen). The eluted phages are used to infect TG1 cells, which are plated on ampicillin/chloramphenicol-containing LB agar plates, and colonies with double resistance ( $Ap^R/Cm^R$ ) are selected.

The phagemids of selected clones are isolated and analyzed by restriction digest (see Figure 7.a) and used as templates for PCR screening. Primer OPEP5L is used to amplify the pYANG3-Ape2, -Ape3 and -Ape10 constructs specifically (see Figure 7.b).

**OPEP5L**     5'- GACTACAAAGATGTCGACTG

**There is a specific enrichment of constructs of correct pairing (Table 2).**

### Example 3: Interactive screening of *E. coli* genomic DNA libraries (Polyphage/tag system)

### 3.1: Principle (see Figure 2)

Instead of using two model libraries as in Example 2, a genomic DNA library of *E. coli* is prepared to be screened against itself to identify interacting *E. coli* peptides or proteins.

### 3.2: Construction of display and expression vectors for genomic DNA

Expression vectors are constructed having a blunt-end restriction site SmaI inserted either in front of His-tag, Strep-tag (Schmidt and Skerra, 1994) or the C-terminal domain of gIII (gIIIc) *via* oligonucleotide cassettes or PCR.

**The self-complementary oligonucleotides OHIS5 & OHIS3, and OSTREP5 & OSTREP3, are used to create ds DNA cassettes encoding the His-tag, and the Strep-tag, respectively.**

OHIS5 5'- AATTCCCGGGCACCACCACCACCACTGATA

**OHIS3**      5'- AGCTTATCAGTGGTGGTGGTGGTGCCCGGGG

OSTREP5 5'- AATCCCCGGGTCTGCTTGGCGTCACCCGCAGTTCGGTGGT-  
TGATA

OSTREP3 5'- AGCTTATCAACCACCGAACTGCGGGTGACGCCAAGCAGACC-  
CGGGG

The cassettes upon phosphorylation and annealing recreate the *EcoRI* and *HindIII* sites. The cassettes are inserted into pIG1 and pIG3 vectors (Ge *et al.*, 1995) cut by the same restriction enzymes. The resulting vectors are pING1-A1, pING3-A1 (for His tag in pIG1 and pIG3 vectors) and pING1-A2, pING3-A2 (for Strep-tag), respectively. The correct vectors are screened for the presence of *XmaI* site (isoschizomer of *SmaI*) and the constructs are confirmed by sequencing. The *XbaI/HindIII* fragments of these vectors are inserted into pBS13 vector, linearized with the same enzymes, resulting in vectors **pING1-C1**, **pING3-C1** and **pING1-C2**, **pING3-C2**, respectively (see Figure 8).

The *gllC* fragment containing the *SmaI* site is generated from PCR amplification of pIG10.3 vector using primers OGIII5 and OGIII3, where OGIII3 anneals 3' of the gene *III* in the vector:

OGIII5 5'- CGGAATTCCCCGGGGAGCAGAAGCTGATC

OGIII3 5'- TTTTCACTTCACAGGTC

Three rounds of PCR are performed with a hot-start: 92°C, 1 min; 46°C, 1 min; 72°C, 1.5 min. This is followed by 30 rounds of: 92°C, 1 min; 50°C, 1 min; 72°C, 1.5 min.

The PCR product is purified (QIAquick) and digested with *EcoRI* and *HindIII*. The fragment is gel-purified (QIAquick) and ligated into pIG10.3. The sequence of the resulting vector, **pONG3-A** (see Figure 8), is confirmed by restriction analysis and by sequencing.

### 3.2: Selection of Interacting Pairs from *E. coli* Genomic DNA via His-tag

Genomic DNA of *E. coli* strain XL-1 Blue (Stratagene) is isolated using the Blood & Cell Culture DNA Maxi kit (Qiagen) and eluted in TE buffer (pH 8.0). 200 µg of the DNA is

taken and sonicated (50 cycles, 270 mA, 0.5 s/stroke). The fragmented DNA (average size: max. 0.7 kB) is blunt-ended by a fill-in reaction with T4 DNA polymerase.

Vectors pING1-C1 and pONG3-A are digested with *EcoRV* and *SmaI*, the vector fragments are gel-purified (Qiagen). The vector fragments are then ligated with the blunt-ended genomic DNA at 16°C overnight. The ligation mixtures are taken to transform TG1 cells.

The pING1-C1 and pONG3-A transformants are scratched from the plate and used to inoculate 2xYT medium containing Cm/glucose or Amp/glucose, respectively. The pING1-C1 culture is infected with helper-phage (VCSM13 or M13k07) and phage particles are isolated. These phage particles are used to infect log-phase cells containing the pONG3-A library. The resulting culture is plated out on large Amp/Cm/glucose plates.

The colonies are scratched from the surface of the plates above and transferred to 2xYT medium containing Amp/Cm. After 30 min shaking at 37°C, the culture is then induced (1 mM IPTG) for 30 min, infected with helper-phage at 37°C for 30 min and shaken at RT overnight.

The phage particles are harvested from the overnight culture and PEG-precipitated. They are selected on immobilized Ni-NTA (Ni-NTA HisSorb Strips, Qiagen). The eluted phages are used to infect log-phase TG1 cells. Selected protein pairs are characterised by determination of their corresponding DNA sequences.

#### **Example 4: Polyphages and Selection of Correct Pairing Interactions via SIP**

##### **4.1: Principle (see Figure 10)**

The purpose of this experiment is to show that from a combination of 2 libraries one can isolate and identify the correct interacting pairs using the SIP (Selectively Infective Phage: Krebber *et al.*, 1995; the term "IMP" used in the experimental section denotes "Infectivity mediating particle" comprising the N-terminal domains of the gene III protein

of filamentous phage) selection system, and recover the information about both interacting partners *via* the formation and selection of polyphage particles. The library members forming interacting pairs with members of the corresponding library are being 'doped' with library members that do not interact with members of the corresponding library, and thus should not give a positive SIP selection.

## 4.2: Construction of vectors

### 4.2.1: fhag1A (see Figure 11)

- a. The phage vector f17/9-hag (Krebber *et al.*, 1995) is digested with EcoRV and XmnI. The 1.1 kb fragment containing the anti-HAG Ab gene is isolated by agarose gel electrophoresis and purified with a Qiagen gel extraction kit. This fragment is ligated into a pre-digested pIG10.3 vector (EcoRV-XmnI). Ligated DNA is transformed into DH5a cells and positive clones are verified by restriction analysis. The recombinant clone is called **pIGhag1A**. All cloning described above and subsequently are according to standard protocols (Sambrook *et al.*, 1989)
- b. The vector f17/9-hag (Krebber *et al.*, 1995) is digested with EcoRV and StuI. The 7.9 kb fragment is isolated and self-ligated to form the vector **fhag2**.
- c. The chloramphenicol resistance gene (CAT) assembled *via* assembly PCR (Ge and Rudolph, 1997) using the the template pACYC (Cardoso and Schwarz, 1992) (Figure 11a shows the functional map and the sequence of the CAT gene) is amplified by the polymerase chain reaction (PCR) with the primers:  
CAT\_BspEI(for): 5' GAATGCTCATCCGGAGTTC  
CAT\_Bsu36I(rev): 5' TTTCAGTGGCCTCAGGCTAGCACCAGGCGTTTAAG
- d. The PCR is done following standard protocols (Sambrook *et al.*, 1989). The amplified product is digested with BspEI and Bsu36I then ligated into pre-digested fhag2 vector (BspEI-Bsu36I; 7.2 kb fragment) to form **fhag2C**.
- e. The vector fhag2C is digested with EcoRI and the ends made blunt by filling-in with Klenow fragment. The flushed vector is self-ligated to form vector **fhag2CdelEcoRI**.

- f. pIGHAG1A is digested with XbaI and HindIII. The 1.3 kb fragment containing the anti-HAG gene fused with the C-terminal domain of filamentous phage pIII protein is isolated and ligated with a pre-digested fhag2CdelEcoRI phage vector (XbaI-HindIII; 6.4 kb) to create the vector **fhag1A**

#### 4.2.2: **fjun1A** (see Figure 12)

- a. The EcoRV site of pIG10.3 is converted to a Sall site by oligonucleotide site-directed mutagenesis (Sambrook *et al.*, 1989) with primer:

Sall9-9primer(rev) 5'CTGAATGTCGACATCTTTGTAGTC3'

The mutated pIG10.3 is called **pIG10.3 Sall**.

- b. The jun leucine-zipper domain from pOK1 (Grammatikoff *et al.*, 1994) is amplified by PCR with the primers:

jun2(for): 5'ACGCGTCGACGCCGGTGGTCGGATCGCCCGG3'

jun2(rev): 5'AATTCGGCACCACCGTGGTTCATGACT3'

- c. The PCR is done following standard protocols (Sambrook *et al.*, 1989). The amplified product is digested with Sall and EcoRI, then ligated into pre-digested pIG10.3Sall vector (Sall-EcoRI) to form the vector **jun-pIG10.3Sall**.

- d. The vector jun-pIG10.3Sall is digested with XbaI and EcoRI. The 0.14 kb fragment is ligated into the pre-digested vector fhag1A (XbaI-EcoRI; 7kb) to form the vector **fjun1A**.

#### 4.2.3: **fjun1B** (see Figure 13)

- a. The DNA encoding the C-terminal domain including the long linker separating it from the amino terminal domain of the filamentous phage pIII (gIII short) is amplified by PCR using pOK1 (Grammatikoff *et al.*, 1994) as template with the primers:

gIII short(for): 5'GCTTCCGGAGAATTCAATGCTGGCGGCGGCTCT3'

gIII short(rev): 5'CCCCCCCAAGCTTATCAAGACTCCTTATTACG3'

- b. The PCR is done following standard protocols (Sambrook *et al.*, 1989). The amplified product is digested with EcoRI and HindIII, then ligated into pre-digested f<sub>hag1A</sub> vector (EcoRI-HindIII) to form the vector f<sub>jun1B</sub>.

#### 4.2.4: f<sub>pep2\_1b</sub>, f<sub>pep3\_1B</sub>, f<sub>pep10\_1b</sub> (see Figure 14)

- a. These constructs are obtained from a peptide library screened against the intracellular domain of p75, the low affinity receptor of NGF, in a SIP experiment.
- b. A peptide library cassette of cyclic peptides with length variants of 6-16 amino acids is prepared from the oligos:

Groprim: 5'-CATGAATTCGGATCCTCC-3'

Gron10: 5'-CTATGGCGCGCCTGTCGACTGT(M)<sub>6-16</sub>TGTGGTGGTGGAGGATC-CGAATTCATG-3'

where M is a mixture of 19 trinucleotide codons (Virnekäs *et al.*, 1994), excluding the one coding for Cys. The length variation is achieved by coupling 6 trinucleotide positions using the standard coupling procedure, and, for the next 10 coupling cycles, by omitting the capping step during DNA synthesis and by diluting the trinucleotide mixture to achieve stepwise coupling yields of 50%.

The oligos are annealed and filled in with the Klenow fragment of DNA polymerase I to form a double-stranded DNA cassette with standard methods (Sambrook *et al.*, 1989). The cassette is digested with Sall-EcoRI, purified with Qiaex DNA gel extraction kit, and ligated to pre-digested f<sub>jun1B</sub> vector (Sall-EcoRI) to form the peptide library. The ligated peptide library is transformed into competent DH5a cells harboring pUC18/IMP-p75 (see below) and plated on Luria Broth (LB) (30 µg/ml chloramphenicol + 100 µg/ml ampicillin) and incubated overnight at ambient temperature.

- c. The Amp<sup>r</sup> Cm<sup>r</sup> colonies are scraped with LB, and 1 ml of suspension is used to inoculate 25 ml LB (30 µg/ml chloramphenicol + 100 µg/ml ampicillin + 1 mM IPTG). The culture is incubated overnight at room temperature.



- d. The supernatant is separated from the cells by centrifugation (10,000 RPM, 10 min., 4°C). 5 ml of 30% PEG/3M NaCl are added to the supernatant and mixed 100 times. After 1 hour on ice, the phage precipitate is collected by centrifugation (10,000 RPM, 10 min., 4°C). The pellet is resuspended in 1 ml TBS buffer. The suspension is filtered with a 0.45 micron filter (Sartorius).
- e. 100 µl of log phase K91 cells (or any male *E. coli* cells (F-pilus containing)) are infected with 10 µl of phage supernatant, plated on LB (30 µg/ml chloramphenicol) and incubated overnight at ambient temperature.
- f. Chloramphenicol-resistant transductants are picked, and overnight cultures are prepared to isolate DNA for sequencing. From the sequencing, fpep2\_1b, fpep3\_1b, fpep10\_1b containing peptides pe2, pe3, and pe10 are identified.  
 pe2: 5'-TGTTTTTTTCGTGGTGGTTTTTTAATCATAATCCTCGTTATTGT-3'  
 (CysPhePheArgGlyGlyPhePheAsnHisAsnProArgTyrCys)  
 pe3: 5'-TGATTGTTTATCATGCTCATTATCTTGTTGCTAAGTGT-3'  
 (CysIleValTyrHisAlaHisTyrLeuValAlaLysCys)  
 pe10: 5'-TGTTCTTATCATCGTCTTTCTACTCGTGTGTTGT-3'  
 (CysSerTyrHisArgLeuSerThrArgValCys)

#### 4.2.5: fNGF1B (see Figure 15)

- a. The DNA encoding the nerve growth factor (NGF) gene is amplified from pXM NGF (Ibanez *et al.*, 1992) as template with the primers:  
 NGF(for): 5'AAAAAAGTCGACTCATCCACCCACCCAGTC3'  
 NGF(rev): 5'AGGAATTCGCCTCTTCTTGACGCTT3'
- b. The PCR is done following standard protocols (Sambrook *et al.*, 1989). The amplified product is digested with Sall and EcoRI, then ligated into pre-digested fjun1B vector (Sall-EcoRI) to form the vector fNGF1B.

#### 4.2.6: pUC19/IMP-HAG (see Figure 16)

- a. The vector f17/9-hag (Krebber *et al.*, 1995) is digested with EcoRI and HindIII. The 1.4 kb fragment containing the gene fusion of the IMP with the HAG peptide, is isolated and cloned into pre-digested pUC19 (EcoRI-HindIII) to form the vector pUC19/IMP-HAG

#### 4.2.7: pUC18/IMP-p75 (see Figure 17)

- a. The intracellular domain of p75 containing the C-terminal 142 amino acids is amplified from the cDNA clone of p75 (Chao *et al.*, 1986) as template with the primers:  
p75(for): 5' GCTGGCCCGTACGACAAGAGGTGGAACAGCTGC  
p75(rev): 5' TCTCGAAGCTTATCACACTGGGGATGTGGC
- b. The PCR is done following standard protocols (Sambrook *et al.*, 1989). The amplified product is digested with BsiWI and HindIII, then ligated into pre-digested pUC19 vector (BsiWI-HindIII) to form the vector pUC19/IMP-p75.
- c. The vector pUC19/IMP-p75 is digested with XbaI and HindIII. The 1 kb fragment is isolated and cloned into the pre-digested pUC18 vector (XbaI-HindIII) to form the vector pUC18/IMP-p75.

#### 4.2.8: pUC18/IMP-IL16 (see Figure 18)

- a. The IL16 gene is amplified from the clone pcDNA3-ILHu1 (M. Baier, Paul Ehrlich Institute, Germany; Baier *et al.*, 1995; Bannert *et al.*, 1996) as template with the primers:  
f1Bsu36Ifor: 5'AGACTGCCTCAGGCCAGCCCGACCTCAACTCC3'  
f3HindIIIrev2: 5'ATATATAAGCTTTTAGGAGTCTCCAGCAGC3'
- b. The PCR is done following standard protocols (Sambrook *et al.*, 1989). The amplified product is digested with Bsu36I and HindIII, then ligated into pre-digested pUC18/IMP-p75 vector (Bsu36I-HindIII) to form the vector pUC18/IMP-IL16.

#### 4.3: In vivo SIP with co-transformation and polyphage

4.3.1: Combining 2 libraries (Library 1 is fused with gIII while Library 2 is fused to the IMP).

10 ng each of fjun1B, fjun1A, fpep3\_1B, fhag1A, fNGF1B with 500 ng each of pUC18/IMP-p75, pUC18/IMP-HAG, pUC18/IMP-IL16 are co-transformed into DH5a cells by electroporation. The cells are plated on Luria Broth (LB) (30 µg/ml chloramphenicol + 100 µg/ml ampicillin) and incubated overnight at ambient temperature.

The Amp<sup>r</sup> Cm<sup>r</sup> colonies are scraped with LB and 1 ml of suspension is used to inoculate 25 ml LB (30 µg/ml chloramphenicol + 100 µg/ml ampicillin + 1 mM IPTG) followed by incubation overnight at room temperature.

4.3.2: In vivo SIP. The supernatant from the cells is separated by centrifugation (10,000 RPM, 10 min., 4°C). 5 ml of 30% PEG/3M NaCl are added to the supernatant and mixed 100 times. After 1 hour on ice, the phage precipitate is collected by centrifugation (10,000 RPM, 10 min., 4°C). The pellet is resuspended in 1 ml TBS buffer, and the suspension is filtered through a 0.45 micron filter (Sartorius).

200 µl of phage supernatant are used to infect 1.8ml of log phase K91 cells (or any male E. coli cells (F-pilus containing)), and the cells are plated on LB (30 µg/ml chloramphenicol + 100 µg/ml ampicillin) and incubated overnight at ambient temperature.

4.3.3: Testing of infectious polyphage DNA patterns and infectivity. Twenty individual Amp<sup>r</sup> Cm<sup>r</sup> colonies are used to inoculate 5 ml LB (30 µg/ml chloramphenicol + 100 µg/ml ampicillin) in each case and incubated at ambient temperature overnight. Plasmid and RF DNA are isolated from each clone with a Qiagen Miniprep DNA kit. Clones are analysed by restriction analysis with restriction enzymes XbaI and HindIII together with appropriate buffers as supplied and instructed by the manufacturer. The restriction

digests are run in a 0.8% TBE agarose gel at constant voltage of 100V for 1.5 hours. The restriction patterns, together with the relative intensity of the bands (because the phage vectors (fjun1B, fjun1A, fpep3\_1B, fNGF1B, fhag1A) have significantly lower copy numbers than the plasmid vectors) allow to identify correctly interacting pairs. For the pair fhag1A+pUC19/IMP-HAG, an XbaI-HindIII digest will yield a 6.5 kb, 3.3 kb, 1.3 kb, and 0.7 kb fragments, while for the pair fpep3\_1B+pUC18/IMP-p75, the same digest will yield 6.3 kb, 2.8 kb, 1kb, and 0.7kb fragments. A problem though is to distinguish the potential non-cognate combinations of fjun1B or fjun1A with pUC18/IMP-p75 because they would give similar patterns as the fpep3\_1B+pUC18/IMP-p75. To further resolve this, the clones containing identical patterns can be re-digested with BamHI-HindIII. The fjun1A or fjun1B in combination with pUC18/IMP-p75 would yield only 4 fragments - 4.1 kb and 2.9 kb, 2.6 kb, 1.2 kb fragments - while the cognate pair fpep3\_1B+pUC18/IMP-p75 will yield 5 fragments - 3.5 kb, 2.9 kb, 2.6 kb, 1.2 kb, 0.5 kb. To further prove that cognate interacting pairs have been selected, the ability of the clones to form selectively-infective phage particles is tested. Only clones with a cognate pair can form infectious phages. The supernatant from the overnight culture of the individual clones is filtered with a 0.45 micron filter (Sartorius). Ten microliters of phage supernatant are mixed with 100 µl of log phase K91 cells (or any male E. coli cells (F-pilus containing)) for 10 minutes at 37°C. The suspension is plated on LB (30 µg/ml chloramphenicol) and incubated overnight at 37°C. The result is shown in Table 3.b.

In summary (see Figure 19), the results from the above example indicate that among 19 clones analyzed, 8/19 have the cognate pair fpep3\_1B+pUC18/IMP-p75 and produce selectively-infective phage; 1/19 has the fhag1A+pUC19/IMP-HAG combination and produces selectively-infective phage.

### **Example 5: Combination of Multiple Libraries into a Single Phagemid Vector through Recombination, Screening via tag system**

#### **5.1: Principle (see Figure 20)**

To be able to retrieve the genetic information for cognate protein pairs selected via a tag fused to one of the partners, two separate libraries in phagemid vectors are constructed containing the *lox* recombination promoting sites and recombined on one phagemid by action of the *cre* recombinase in an *in vivo* recombination.

### 5.2: Vector construction

Both *loxP* and *loxP511* sites (Hoess *et al.*, 1986) are inserted in tandem into the region flanked by the ColE1 ori and  $\beta$ -lactamase in vector pING1-C1, whereas in vector pONG3-A, the *loxP* site is cloned upstream of the *Xba*I site and the *loxP511* downstream of the *Hind*III site. Therefore, the genomic DNAs to be cloned are flanked by the *loxP* and *loxP511* sites.

### 5.3: Library construction and recombination

The libraries are prepared as in Example 3. The phagemids in the double-resistant clones are recombined through the *cre* recombinase which either is encoded in the phagemid being inducible (Tsurushita *et al.*, 1996), or is transferred through P1 phage infection (Rosner, 1972; Waterhouse *et al.*, 1993). Phages are prepared from the recombined clones by helper phage infection and used to infect new *E. coli* cells (*cre*<sup>-</sup>).

### 5.4: Selection

The phage particles are prepared from the Cm<sup>R</sup> clones and subjected to His-tag selection as in Examples 2 and 3. The sequences encoded in each phagemid, which now contains members of both libraries, can be determined by sequencing using primers specific for myc-tag region (library 1) and His-tag region (library 2).

**Example 6: SIP-based library vs. library screening via *in vitro* recombination of separately constructed libraries into one phage vector**

### 6.1: Principle (see Figure 21)

To be able to retrieve the genetic information for cognate protein pairs selected by SIP interaction *in vivo*, two separate libraries in phage and plasmid vectors are constructed and recombined by co-ligation in an *in vitro* recombination.

### 6.2: Construction of Libraries A and B

**Library A** encodes 2 members, namely a single chain Fv antibody against a peptide derived from hemagglutinin ( $\alpha$ hag) and the leucine zipper domain derived from the jun transcription factor (fjun), both N-terminally fused to the C-terminal domain of gIII from filamentous phage and preceded by the ompA signal sequence followed by the Flag epitope.

**Library B** encodes 3 members on plasmid vectors of the pUC series, namely the hemagglutinin peptide to which the above  $\alpha$ hag antibody binds (pUC19-IMP<sub>hag</sub>), the leucine zipper domain of the fos transcription factor (pUC18-IMP<sub>fos</sub>) which heterodimerizes with jun via this domain, and the intracellular domain of the low affinity nerve growth factor receptor (pUC18-IMP<sub>p75</sub>), as a negative control which does not interact with library A members, all fused to the infectivity-mediating N-terminal domains of phage gIII protein, preceded by the gIII signal sequence.

Library A members are cloned into a fd phage vector which also contains downstream of the library A insertion site the N-terminal domains (N1-N2) of gIII, followed by the cloning sites *Bs*WI and *Hind*III to allow in-frame insertion of library B members.

Library A construct  $\alpha$ hag is identical to the f17/9-hag fd phage vector (Krebber *et al.*, 1995) and serves as basis for construction of fjun. The jun leucine zipper together with amino acids 290 to 326 of the C-terminal part of gIII is PCR-amplified (primers FR620 and FR621, containing *Eco*RV and *Sfi*I sites, respectively) from the construct fjun1B (containing the jun leucine zipper fused to amino acids 290 to 493 of gIII) generated in Example 4. The resulting PCR fragment is ligated directionally into *Eco*RV/*Sfi*I-digested

f17/9-hag vector in frame with amino acids 327 to 493 of the gIII C-terminal domain resulting in vector **fjunhag** (see Figure 22).

Generation of library B constructs pUC19-IMPhag and pUC18-IMPp75 is described in Example 4. To construct pUC18-IMPfos, amino acids 219 to 272 of the N-terminal part of gIII together with the fos leucine zipper are PCR-amplified (primers FR618 and FR619, containing BsiWI and HindIII sites, respectively) from the pOK1 phagemid vector (Grammatikoff *et al.*, 1994). The resulting PCR fragment is ligated directionally into BsiWI/HindIII-digested pUC18-IMPp75 to create **pUC18-IMPfos** (see Figure 17).

**Primers:**

FR618: 5'CGCCGTACGGCGGCTCTGGTGGTGGTTCTGGTGGC3'

FR619: 5'CCCAAGCTTTTAGACTAGCTGACTAGAAGATCTGC3'

FR620: 5'CGCGATATCGTCGACGCCGGTGGTCGGATCGCC3'

FR621: 5'CGCGGCCCCCGAGGCCCCACCACCGGAACCGCCTCCC3'

### 6.3: Preparation and recombination of library A and B and selection of interacting protein pairs by SIP

Non-covalent, cognate interactions of  $\alpha$ hag antibody with hag peptide (Krebber *et al.* 1995) and of fos and jun leucine zipper domains (Grammatikoff *et al.*, 1994) generates infective SIP phage. Thus, from the six possible combinations of members of the model libraries A and B (**fahag-hag**, fahag-fos, fahag-p75, **fjun-fos**, fjun-hag, fjun-p75), only two combinations (cognate pairs in bold) should be selected by *in vivo* SIP. To recombine the library members in all possible permutations, library A is linearized by digestion with BsiWI/HindIII to prepare it for random incorporation of library B members, prepared by mass-excision with BsiWI/HindIII from the construct B pool described above. After co-ligation of the mass-excised library B fragments into library A vectors, the sample is transformed into competent E.coli cells, plated onto chloramphenicol-containing LB agar plates and grown overnight at 37°C. The recombined library size can be determined by plating serial dilutions of the transformation and can be compared to

the complexities of the individual libraries A and B. The total recombined library is scraped from the plates in LB medium and used to inoculate an appropriate volume of chloramphenicol-selective LB-medium supplemented with 1 mM IPTG. After growth at 30°C overnight with constant shaking to allow production of SIP phages, the bacteria are pelleted by centrifugation and phages present in the supernatant are precipitated on ice for one hour by addition of 0.25 volumes of 20% PEG/2.5 M NaCl. The phages are pelleted by centrifugation for 30 min at 10 000 x g and 4°C. The pellet is resuspended in an appropriate volume of 1 x TBS buffer and filtered through a 0.45 µm filter. Serial dilutions of this filtrate are used to infect F<sup>+</sup> E.coli cells. The double-stranded, replicative form phage DNA is prepared from resulting transductant colonies by standard methods and analyzed by restriction digest and sequencing for the presence and identity of library A and B members. Furthermore, the supernatant of transductant colonies is analyzed for the presence of infective SIP phages to confirm that protein-protein interaction of a particular pair selected from the recombined libraries A and B is responsible for SIP phage infectivity.

Alternatively, the model libraries A (2 members) and B (3 members) are used to construct all possible combinations (listed above) individually, and equal amounts (50 ng) of each of the 6 combinations can be co-transformed into competent E. coli cells followed by the steps listed above. The distribution of individual constructs after co-transformation as well as the distribution of transductants resulting from the model library can be analyzed as described above. The selective recovery of phage constructs which co-encode cognate protein pairs demonstrates the feasibility of SIP-based selection of binding partners after an appropriate recombination event.

## **Example 7: 'Spatial' *in vivo* SIP**

### **7.1: Principle (see Figure 23)**

Coupling of information about members of interacting peptides or proteins is achieved by having a spatial relationship between the particles displaying the selectable or



screenable property (in this example phages for the SIP experiment) and the package containing the genetic information for the individual library members (in this example the *E. coli* cell secreting the phage particle being screened), i. e. a correlation between the phage being examined and the position of the corresponding *E. coli* host on the master plate.

#### **7.2: Combining 2 libraries (Library A is fused with gIII while library B is fused to the IMP)**

10 ng each of fjun1B, fjun1A, fpep3\_1B, fhag1A, fNGF1B are co-transformed with 500 ng each of pUC18/IMP-p75, pUC19/IMP-HAG, pUC18/IMP-IL16 into DH5a cells by electroporation. The transformants are plated on LB (30 µg/ml chloramphenicol + 100 µg/ml ampicillin) and incubated overnight at ambient temperature.

#### **7.3: Screening of co-transformants by SIP**

From the master plate of co-transformants, each of the co-transformants are labelled and inoculated separately into 5 ml LB (30 µg/ml chloramphenicol + 100 µg/ml ampicillin) and incubated overnight at ambient temperature.

Plasmid and RF DNA are isolated from each clones with a Qiagen Miniprep DNA kit. Clones are analysed by restriction analysis with restriction enzymes XbaI and HindIII together with appropriate buffers as supplied and instructed by the manufacturer. The restriction digests are run in a 0.8% TBE agarose gel at constant voltage of 100 V for 1 to 2 hours. Restriction patterns allow discrimination of the particular clones.

The supernatant from the overnight culture of the individual clones is filtered with a 0.45 micron filter (Sartorius). Ten microliters of phage supernatant are mixed with 100 µl of log phase K91 cells (or any male *E. coli* cells (F-pilus containing)) for 10 minutes at 37°C. The suspension is plated on LB (30 µg/ml chloramphenicol) and incubated overnight at 37°C.

A positive co-transformant (i.e. contains the correct interacting pair) has a corresponding correct restriction pattern and is capable of producing infectious phages, that are incapable of secondary or subsequent infections. Polyphage particles being capable of such infections, and containing the genetic information of an interacting pair as well, can readily be identified by their restriction digest pattern.

### **Example 8: *E. coli* display**

#### **8.1: Principle (see Figure 24)**

Two libraries are introduced into *E. coli* cells, with expressed members of library A (such as antibody, peptide, or cDNA libraries) being presented at the surface of the cells. In those cases where interacting pairs are formed, members of library B (such as antibody, peptide, or cDNA libraries) are transported in the complex with its cognate partner to the surface of the cell as well, thus displaying a selectable or screenable property such as a tag. Selected cells contain the information for both interacting partners.

#### **8.2: Preparation of Library A**

A thioredoxin peptide library is prepared as fusions to the *E. coli* flagellin in the pFLITRX vector essentially as described (Lu *et al.*, 1995).

#### **8.3: Preparation of Library B**

An cyclic, variable-length peptide library including a FLAG epitope (Hopp *et al.*, 1988; Knappik and Plückthun, 1994) is prepared essentially as described in Example 4.2.4, and cloned in the pTERM vector, a modified version of the pto2H10a3s vector (Krebber *et al.*, 1996) containing a chloramphenicol-resistance gene instead of the kanamycin-resistance gene. The pTERM vector can be assembled by standard methods starting from pto2H10a3s. This cyclic peptide library is packaged by infection with a helper phage (M13K07 or VCSM13) by standard methods (Sambrook *et al.*, 1989).

#### 8.4: Combination of Library A and Library B

An aliquot of the *E. coli* cells containing Library A is used to inoculate 50 ml LB (100 µg/ml ampicillin) and incubated at ambient temperature until the OD600 reached 0.4. The cells are infected with phages containing Library B at a multiplicity of infection (MOI) of 10. After 30 min of infection, the cells are collected by centrifugation (5000 RPM, 10 minutes, 4°C) and resuspended in 1 ml LB. The suspension is plated on M9 media (+ 1 mM MgCl<sub>2</sub>, supplemented with 0.5% glucose, 0.2% casamino acids, 100 µg/ml ampicillin, 30 µg/ml chloramphenicol).

#### 8.5: Selection of Interacting pairs

The Amp<sup>r</sup> Cm<sup>r</sup> colonies are scraped with M9 media (+ 1 mM MgCl<sub>2</sub>, supplemented with 0.5% glucose, 0.2% casamino acids, 100 µg/ml ampicillin, 30 µg/ml chloramphenicol), and an aliquot of the suspension is used to inoculate 25 ml M9 media (+ 1 mM MgCl<sub>2</sub>, supplemented with 0.5% glucose, 0.2% casamino acids, 100 µg/ml ampicillin, 30 µg/ml chloramphenicol) and incubated at 37°C until saturation. Selection is performed essentially as described (Lu *et al.*, 1995), the modification being that the antibody used for selection is the M1 anti-FLAG antibody (Kodak).

Individual enriched Amp<sup>r</sup> Cm<sup>r</sup> colonies are isolated and the sequences of the corresponding interacting peptide(s) and cyclic peptide(s) are determined by DNA sequencing. To confirm that the encoded peptide and cyclic peptide form a cognate pair, each of the clones is tested for enrichment based on the selection method described above, whereby the Amp<sup>r</sup> Cm<sup>r</sup> colonies bind to the M1 anti-FLAG antibody in a single round of selection.

#### Literature:

Baier *et al.*, 1995, *Nature* 378, 563

Bannert *et al.*, 1996, *Nature* 381, 30

Cardoso and Schwarz, *J. Appl. Bacteriol.* 72 (1992) 289-293

- Cramer and Suter, 1993, *Gene* 137, 69-75
- Ge *et al.*, 1995, *Antibody Engineering*, 2<sup>nd</sup> ed., 229-266
- Ge and Rudolph, *BioTechniques* 22 (1997) 28-29
- Gramatikoff *et al.*, *Nucleic Acids Res.* 22 (1994) 5761-5762
- Hochuli *et al.*, 1988, *Bio/Technology* 6, 1321-1325
- Hoess *et al.* 1986, *Nucleic Acids Res.* 14, 2287-2300
- Hopp *et al.*, 1988, *Bio/Technology* 6, 1204-1210
- Ibanez *et al.*, 1992, *Cell* 69, 329-341
- Knappik and Plückthun, 1994, *BioTechniques* 17, 754-761
- Krebber *et al.*, 1995, *FEBS Letters* 377, 227-231
- Krebber *et al.*, 1996, *Gene* 178, 71-74
- Lindner *et al.*, 1992, *Methods: A companion to Methods Enzymol.* 4, 41-56
- Lu *et al.*, 1995, *Bio/Technology* 13, 366-372
- Rosner, 1972, *Virology* 48, 679-689
- Sambrook *et al.*, 1989, *Molecular Cloning: a Laboratory Manual*, 2<sup>nd</sup> ed.
- Schmidt and Skerra, 1994, *J. Chromatogr. A* 676, 337-345
- Tsurushita *et al.*, 1996, *Gene* 172, 59-63
- Vimekäs *et al.*, 1994, *Nucleic Acids Res.* 22, 5600-5607
- Waterhouse *et al.*, 1993, *Nucleic Acids Res.* 21, 2265-2266

## CLAIMS

1. A method for identifying a plurality of nucleic acid sequences, said nucleic acid sequences each encoding a (poly)peptide capable of interacting with at least one further (poly)peptide encoded by a different member of said plurality of nucleic acid sequences, comprising the steps of:
  - (a) providing a first library of recombinant vector molecules containing genetically diverse nucleic acid sequences comprising a variety of nucleic acid sequences encoding (poly)peptides;
  - (b) providing a second library of recombinant vector molecules containing genetically diverse nucleic acid sequences comprising a variety of nucleic acid sequences encoding (poly)peptides capable of interacting with further (poly)peptides as mentioned in step (a), wherein the vector molecules employed for the production of said recombinant vector molecules and/or the recombinant inserts display properties that are phenotypically distinguishable from those of the vector molecules and/or the recombinant inserts used in step (a) and wherein at least one of said properties displayed by each of said vector molecules and/or the recombinant inserts used in steps (a) and (b), upon the interaction of a (poly)peptide from said first library with a (poly)peptide from said second library together generate a screenable or selectable property;
  - (c) optionally, providing additional libraries of recombinant vector molecules containing genetically diverse nucleic acid sequences comprising a variety of nucleic acid sequences encoding (poly)peptides capable of interacting with or causing interaction of (a) further (poly)peptide(s) as mentioned in step (a) and/or step (b), wherein the vector molecules employed for the

production of said recombinant vector molecules and/or the recombinant inserts display properties that are phenotypically distinguishable from those of the vector molecules and/or the recombinant inserts used in steps (a) and (b) and, optionally, at least one of said properties displayed by said vector molecule and/or the recombinant inserts used in step (c) together with at least one of said properties displayed by either said vector molecule and/or said recombinant inserts used in steps (a) and/or (b), upon the interaction of a (poly)peptide from said additional library with either a (poly)peptide from said first library and/or a (poly)peptide from said second library generate a screenable or selectable property;

- (d) expressing members of said libraries of recombinant vectors or nucleic acid sequences mentioned in steps (a), (b) and optionally (c), in appropriate host cells so that at least one interaction is established;
- (e) selecting for the generation of said screenable or selectable property representing the interaction of said (poly)peptides;
- (f) optionally, carrying out further screening, selection and/or purification steps; and
- (g) identifying said nucleic acid sequences encoding said (poly)peptides.

2. A method for identifying a plurality of nucleic acid sequences, said nucleic acid sequences each encoding a (poly)peptide capable of interacting with at least one further (poly)peptide encoded by a different member of said plurality of nucleic acid sequences, comprising the steps of:

- (a) expressing in appropriate host cells
  - (aa) nucleic acid sequences contained in a first library of recombinant vector molecules containing genetically diverse nucleic acid sequences comprising a variety of nucleic acid sequences encoding (poly)peptides;

- (ab) nucleic acid sequences contained in a second library of recombinant vector molecules containing genetically diverse nucleic acid sequences comprising a variety of nucleic acid sequences encoding (poly)peptides capable of interacting with further (poly)peptides as mentioned in step (aa), wherein the vector molecules employed for the production of said recombinant vector molecules and/or the recombinant inserts display properties that are phenotypically distinguishable from those of the vector molecules and/or the recombinant inserts used in step (aa) and wherein at least one of said properties displayed by each of said vector molecules and/or the recombinant inserts used in steps (aa) and (ab), upon the interaction of a (poly)peptide from said first library with a (poly)peptide from said second library together generate a screenable or selectable property;
- (ac) optionally, nucleic acid sequences contained in additional libraries of recombinant vector molecules containing genetically diverse nucleic acid sequences comprising a variety of nucleic acid sequences encoding (poly)peptides capable of interacting with or causing interaction of (a) further (poly)peptide(s) as mentioned in step (aa) and/or step (ab), wherein the vector molecules employed for the production of said recombinant vector molecules and/or the recombinant inserts display properties that are phenotypically distinguishable from those of the vector molecules and/or the recombinant inserts used in steps (aa) and (ab) and, optionally, at least one of said properties displayed by said vector molecule and/or the recombinant inserts used in step (ac) together with at least one of said properties displayed by either said vector molecule and/or said recombinant inserts used in steps (aa) and/or (ab), upon the

interaction of a (poly)peptide from said additional library with either a (poly)peptide from said first library and/or a (poly)peptide from said second library generate a screenable or selectable property;

so that at least one interaction is established;

- (b) selecting for the generation of said screenable or selectable property representing the interaction of said (poly)peptides;
  - (c) optionally, carrying out further screening, selection and/or purification steps; and
  - (d) identifying said nucleic acid sequences encoding said (poly)peptides.
3. The method according to claim 1 or 2, wherein said screenable or selectable property is expressed extracellularly.
4. The method according to any one of claims 1 to 3 wherein said recombinant vector molecules in step (a)/(aa) give rise to a replicable genetic package (RGP) displaying said (poly)peptides at its surface.
5. The method according to claim 4, wherein said recombinant vector molecule is a recombinant phage, phagemid or virus.
6. The method according to claim 5, wherein said phage is
- (a) one of the class I phage fd, M13, If, Ike, ZJ/2, Ff;
  - (b) one of the class II phage Xf, Pf1, and Pf3;
  - (c) one of the lambdoid phages, lambda, 434, P1;
  - (d) one of the class of enveloped phages, PRD1; or
  - (e) one of the class paramyxo-viruses, orthomyxo-viruses, baculo-viruses, retro-viruses, reo-viruses and alpha-viruses.



7. The method according to any one of claims 4 to 6, wherein said selection step (e)/(b) is carried out by selecting polyphage comprising the interacting (poly)peptides.
8. The method according to any one of claims 4 to 7, wherein said screenable or selectable property is connected to the infectivity of said RGP.
9. The method according to claim 8, wherein said RGP is encoded by said recombinant vector used in step (a)/(aa) and rendered non-infective and infectivity of said RGP is restored by interaction of said (poly)peptide of step (a)/(aa) with the (poly)peptide of step (b)/(ab) and/or (c)/(ac), said (poly)peptide of step (b)/(ab) and/or (c)/(ac) being fused to a domain that confers infectivity to said RGP.
10. The method according to claim 9, wherein said RGP is rendered non-infective by modification of a genetic sequence which encodes a surface protein necessary for the RGP's binding to and infection of a host cell.
11. The method according to any one of claims 1 to 3, wherein said recombinant vector molecules in step (a)/(aa) give rise to a fusion protein which is expressed on the surface of a cell, preferably a bacterium.
12. The method according to claim 11, wherein said bacterium is *Neisseria gonorrhoe* or *E. coli* and said fusion protein consists of at least a part of a flagellum, lam B, peptidoglycan-associated lipoprotein or the Omp A protein and said (poly)peptide.

13. The method according to any one of claims 3 to 7, 11 or 12, wherein said (poly)peptides encoded by said recombinant vector molecules of step (b)/(ab) or (c)/(ac) are linked to at least one screenable or selectable tag.
14. The method according to claim 13, wherein said screenable or selectable tag is encoded by said recombinant vector of step (b)/(ab) or (c)/(ac).
15. The method according to claim 13 or 14, wherein said screenable or selectable tag is selected from the list His(n), myc, FLAG, malE, thioredoxin, GST, streptavidin, beta-galactosidase, alkaline phosphatase, T7 gene 10, Strep-tag and calmodulin.
16. The method according to claim 13, wherein said screenable or selectable tag is encoded by the genome of the host cell.
17. The method according to any one of claims 1 to 16, wherein said (poly)peptides encoded by the nucleic acid sequences of said additional libraries of step (c)/(ac) cause the interaction of said (poly)peptides of steps (a)/(aa) and (b)/(ab) via phosphorylation, glycosylation, methylation, lipidation or farnesylation of at least one of said (poly)peptides of steps (a)/(aa) and (b)/(ab).
18. The method according to any of claims 1 to 10 and 13 to 17, wherein said host cells in step (d)/(a) are spatially addressable, and the nucleic acid sequences mentioned in step (g)/(d) are retrieved from the corresponding spatially addressable host cell.
19. The method according to claim 1 or 2, wherein said screenable or selectable property is expressed intracellularly.

20. The method according to claim 19, wherein said screenable or selectable property is the transactivation of transcription of a reporter gene such as beta-galactosidase, alkaline phosphatase or nutritional markers such as his3 and leu, or resistance genes giving resistance to an antibiotic such as ampicillin, chloramphenicol, kanamycin, zeocin, neomycin, tetracycline or streptomycin.
21. The method according to any one of claims 1 to 20, wherein said recombinant vectors of step (a)/(aa), (b)/(ab) and (c)/(ac) comprise recombination promoting sites and in said step (e)/(b) recombination events are selected for, wherein said nucleic acid sequences encoding said (poly)peptides of step (a)/(aa), said nucleic acid sequences encoding said (poly)peptides of step (b)/(ab) and optionally said nucleic acid sequences encoding said (poly)peptides of step (c)/(ac) are contained in the same vector.
22. The method according to claim 21, wherein said recombination events are mediated by the site-specific recombination mechanisms Cre-lox, attP-attB, Mu gin or yeast flp.
23. The method according to claim 21 wherein said recombination promotion sites are restriction enzyme recognition sites and said recombination event is achieved by cutting the recombinant vector molecules mentioned in step (a)/(aa), (b)/(ab) and optionally (c)/(ac) with at least two different restriction enzymes and effecting recombination of the nucleic acid sequences contained in said vectors by ligation.
24. The method according to any one of claims 1 to 23 wherein said identification of said nucleic acid sequences is effected after the selection of step (e)/(b) via PCR and preferably sequencing of said nucleic acid sequences after said PCR.

25. The method according to any one of claims 1 to 24, wherein said recombinant vectors of step (a)/(aa), (b)/(ab) and/or (c)/(ac) comprise at least one gene encoding a selection marker.
26. The method according to claim 25, wherein said selection marker is a resistance to an antibiotic, preferably to ampicillin, chloramphenicol, kanamycin, zeocin, neomycin, tetracycline or streptomycin.
27. The method according to any one of claims 1 to 26, wherein said host cells are F' and preferably E.coli XL-1 Blue, K91 or its derivatives, TG1, XL1kan or TOP10F.
28. The method according to any one of claims 3 to 18 and 21 to 27, wherein said RGP's are produced with the use of helper phage taken from the list R408, M13k07 and VCSM13, M13de13, fCA55 and fKN16 or derivatives thereof.
29. The method according to any of claims 1 to 28, wherein at least one of said genetically diverse nucleic acid sequences encode members of the immunoglobulin superfamily.
30. The method according to claim 29, wherein said genetically diverse nucleic acid sequences encode a repertoire of immunoglobulin heavy or light chains.
31. The method according to any of claims 1 to 30, in which said genetically diverse nucleic acid sequences are generated by a mutagenesis method.
32. The method according to any of claims 1 to 31, in which said genetically diverse nucleic acid sequences are generated from a cDNA library.

33. The method according to any one of claims 1 to 32 wherein said nucleic acid sequences are genes or parts thereof.
34. Kit comprising at least
- (a) a recombinant vector molecule as described in step (a)/(aa) or a corresponding vector molecule;
  - (b) a recombinant vector molecule as described in step (b)/(ab) or a corresponding vector molecule; and, optionally,
  - (c) at least one further recombinant vector molecule as described in step (c)/(ac) or a corresponding vector molecule.

Figure 1: General description of the polyphage principle

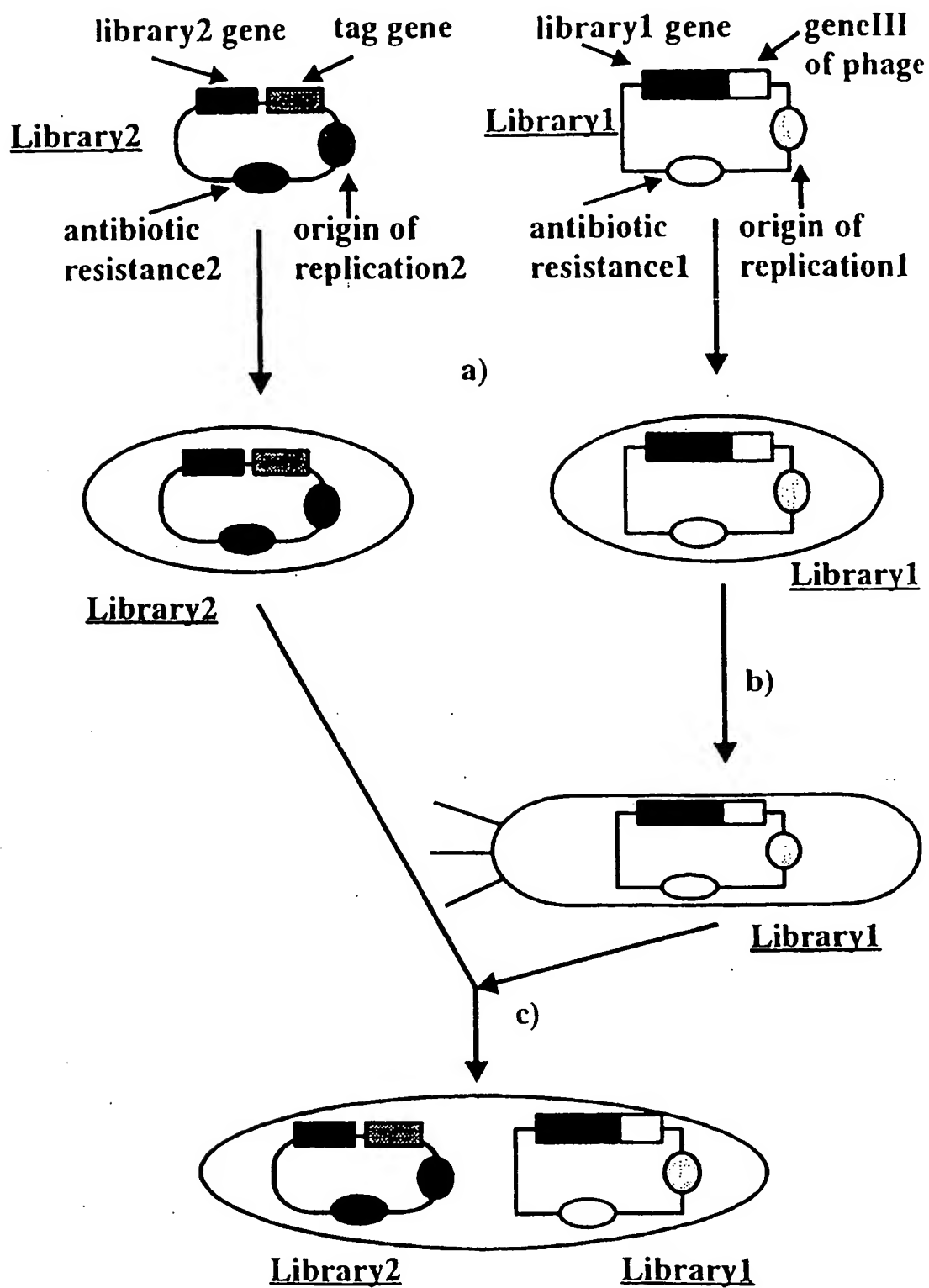


Figure 1: General description of the polyphage principle (cont.)

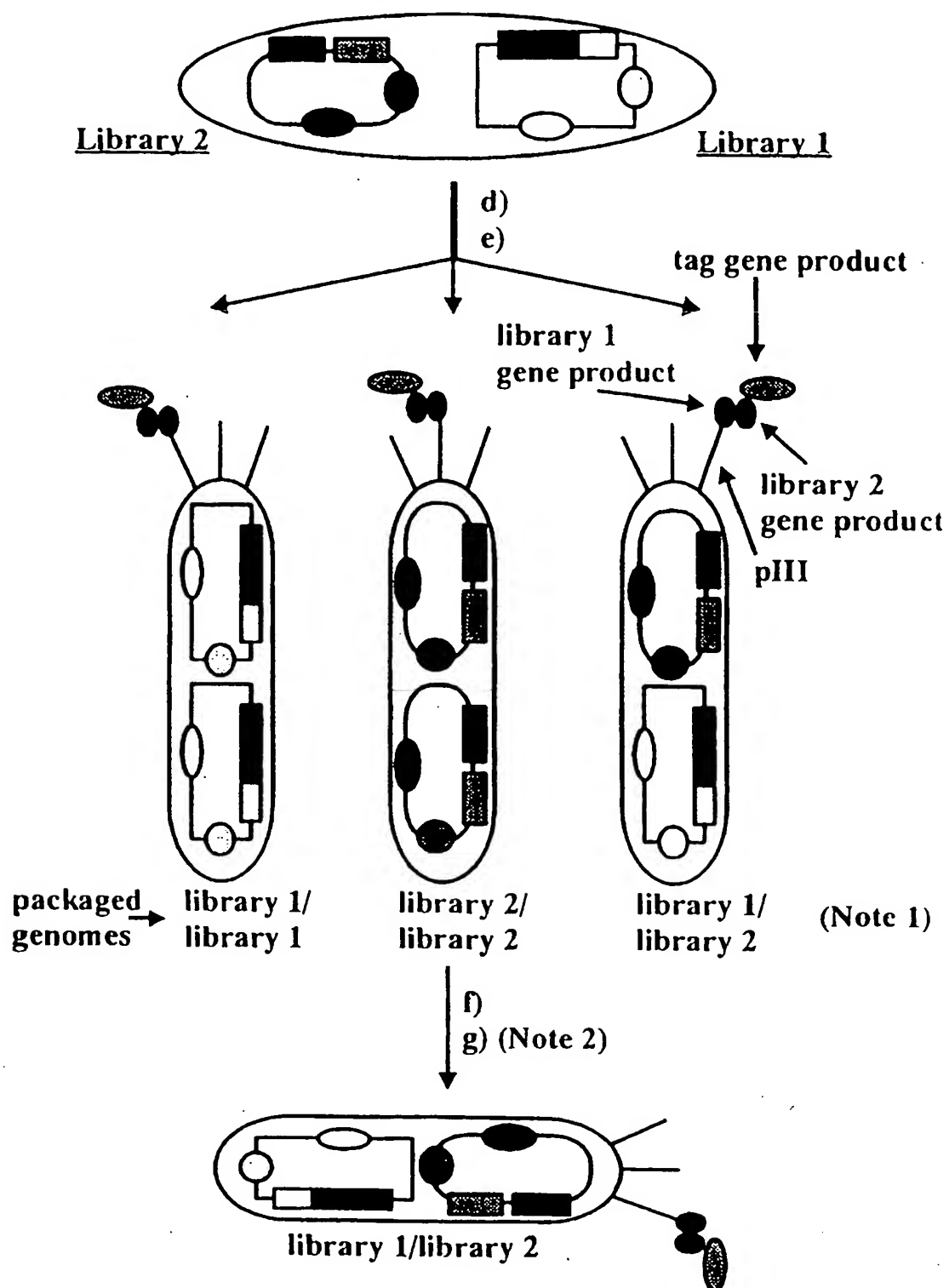


Figure 2: Co-transformation of two phagemids, polyphage formation and selection *via* His-tag: general description

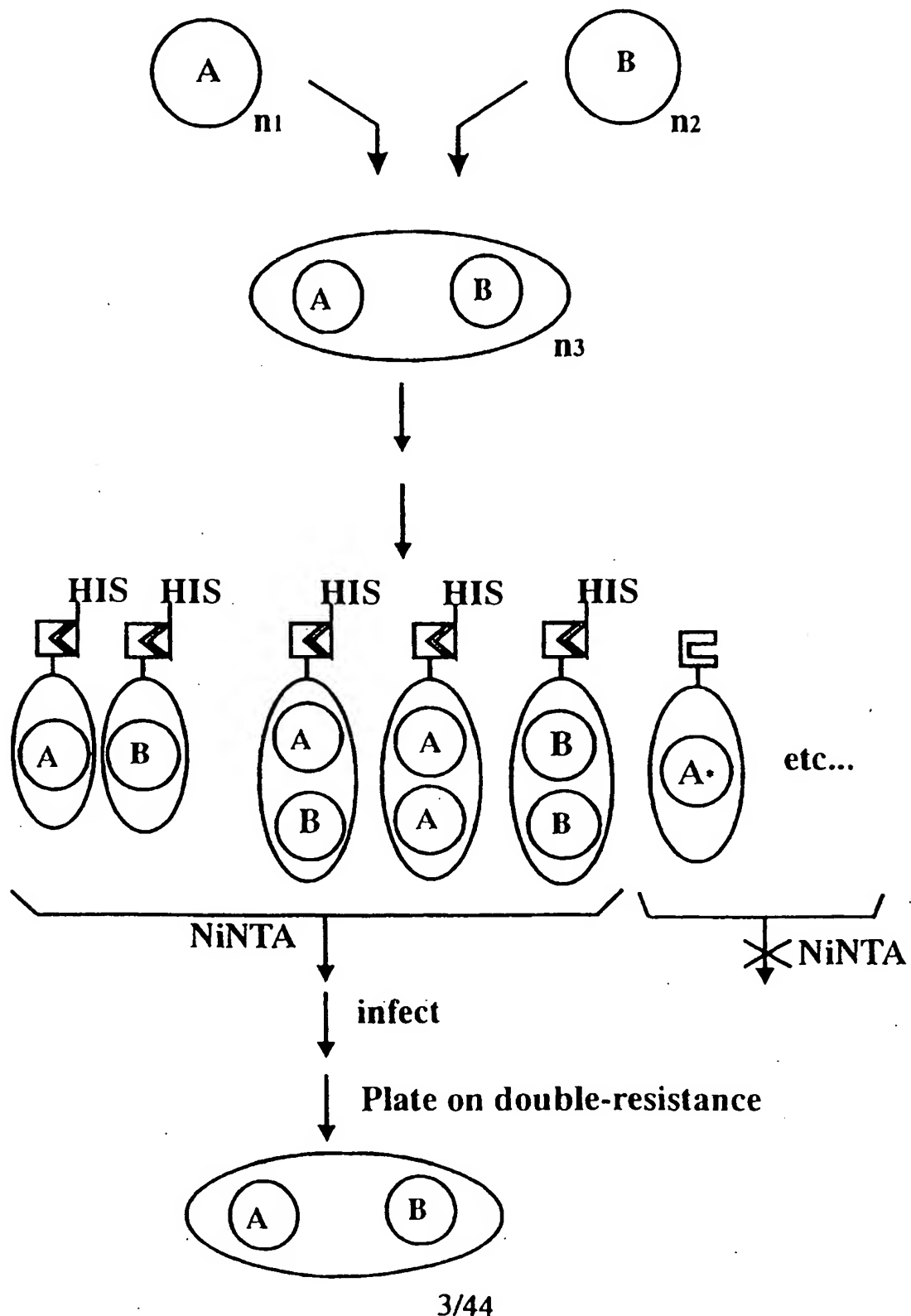




Figure 3: pBS vector series: functional map and sequence of pBS13

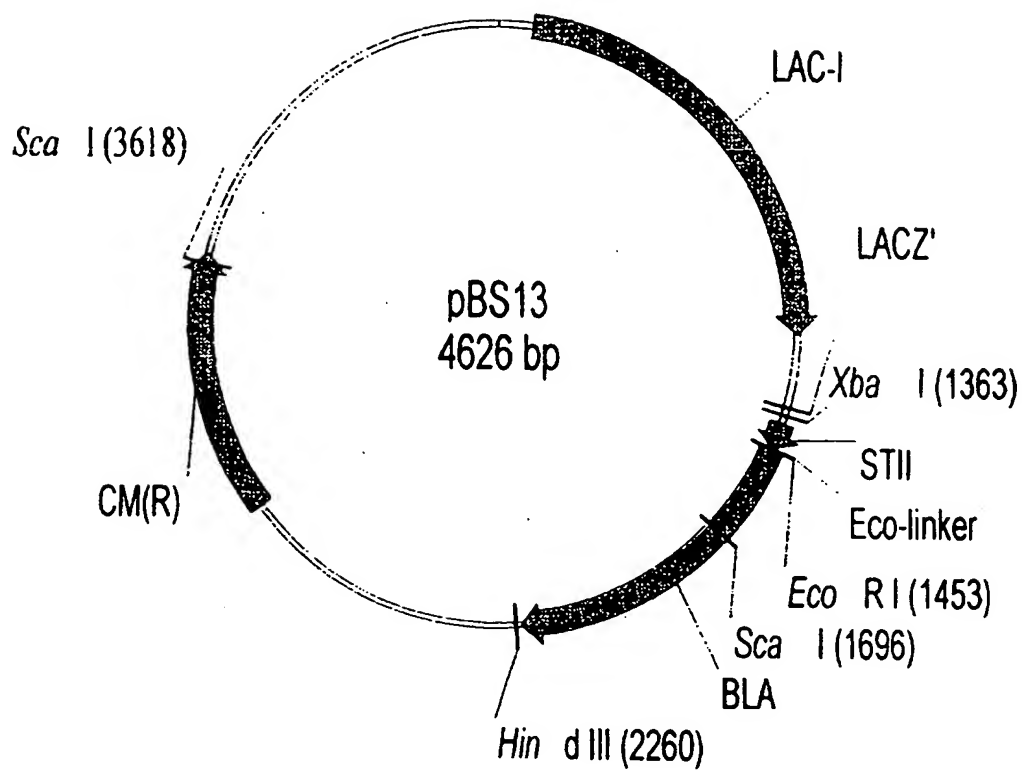


Figure 3: pBS vector series: functional map and sequence of pBS13 (continued)

```

1  ACCCGACACC ATCGAATGGC GCAAAACCTT TCGCGGTATG GCATGATAGC
   TGGGCTGTGG TAGCTTACCG CGTTTTGGAA AGCGCCATAC CGTACTATCG

51  GCCCGGAAGA GAGTCAATTC AGGGTGGTGA ATGTGAAACC AGTAACGTTA
   CGGGCCTTCT CTCAGTTAAG TCCCACCACT TACACTTTGG TCATTGCAAT

101 TACGATGTCG CAGAGTATGC CGGTGTCTCT TATCAGACCG TTTCCCGCGT
   ATGCTACAGC GTCTCATACG GCCACAGAGA ATAGTCTGGC AAAGGGCGCA

151 GGTGAACCAG GCCAGCCACG TTTCTGCGAA AACGCGGGAA AAAGTGGAAAG
   CCACTTGGTC CGGTCGGTGC AAAGACGCTT TTGCGCCCTT TTTCACCTTC

201 CGGCGATGGC GGAGCTGAAT TACATTCCCA ACCGCGTGGC ACAACAAC TG
   GCCGCTACCG CCTCGACTTA ATGTAAGGGT TGGCGCACCG TGTGTGTTGAC

251 GCGGGCAAAC AGTCGTTGCT GATTGGCGTT GCCACCTCCA GTCTGGCCCT
   CGCCCGTTTG TCAGCAACGA CTAACCGCAA CGGTGGAGGT CAGACCGGGA

301 GCACGCGCCG TCGCAAATTG TCGCGGCGAT TAAATCTCGC GCCGATCAAC
   CGTGCGCGGC AGCGTTTAAC AGCGCCGCTA ATTTAGAGCG CGGCTAGTTG

351 TGGGTGCCAG CGTGGTGGTG TCGATGGTAG AACGAAGCGG CGTCGAAGCC
   ACCCACGGTC GCACCACCAC AGCTACCATC TTGCTTCGCC GCAGCTTCGG

401 TGTAAGCGG CGGTGCACAA TCTTCTCGCG CAACGCGTCA GTGGGCTGAT
   ACATTTGCCG GCCACGTGTT AGAAGAGCGC GTTGCGCAGT CACCCGACTA

451 CATTAAC TAT CCGCTGGATG ACCAGGATGC CATTGCTGTG GAAGCTGCCT
   GTAATTGATA GGC GACCTAC TGGTCCTACG GTAACGACAC CTTGACGGA

501 GCACTAATGT TCCGGCGTTA TTTCTTGATG TCTCTGACCA GACACCCATC
   CGTGATTACA AGGCCGCAAT AAAGAACTAC AGAGACTGGT CTGTGGGTAG

551 AACAGTATTA TTTTCTCCCA TGAAGACGGT ACGCGACTGG GCGTGGAGCA
   TTGTCATAAT AAAAGAGGGT ACTTCTGCCA TGC GCTGACC CGCACCTCGT

601 TCTGGTCGCA TTGGGTCACC AGCAAATCGC GCTGTTAGCG GGCCCATTA
   AGACCAGCGT AACCCAGTGG TCGTTTAGCG CGACAATCGC CCGGGTAATT

651 GTTCTGTCTC GCGCGTCTG CGTCTGGCTG GCTGGCATAA ATATCTCACT
   CAAGACAGAG CCGCGCAGAC GCAGACCGAC CGACCGTATT TATAGAGTGA

701 CGCAATCAAA TTCAGCCGAT AGCGGAACGG GAAGGCGACT GGAGTGCCAT
   GCGTTAGTTT AAGTCGGCTA TCGCCTTGCC CTTCCGCTGA CCTCACGGTA

751 GTCCGGTTTT CAACAAACCA TGCAAATGCT GAATGAGGGC ATCGTTCCCA
   CAGGCCAAAA GTTGTGTTGGT ACGTTTACGA CTTACTCCCG TAGCAAGGGT

801 CTGCGATGCT GGTGCGCAAC GATCAGATGG CGCTGGGCGC AATGCGCGCC
   GACGCTACGA CCAACGGTTG CTAGTCTACC GCGACCCGCG TTACGCGCGG

```

Figure 3: pBS vector series: functional map and sequence of pBS13 (continued)

```

851  ATTACCGAGT CCGGGCTGCG CGTTGGTGCG GACATCTCGG TAGTGGGATA
    TAATGGCTCA GGCCCGACGC GCAACCACGC CTGTAGAGCC ATCACCCCTAT

901  CGACGATACC GAAGACAGCT CATGTTATAT CCCGCCGTTA ACCACCATCA
    GCTGCTATGG CTTCTGTCGA GTACAATATA GGGCGGCAAT TGGTGGTAGT

951  AACAGGATTT TCGCCTGCTG GGGCAAACCA GCGTGGACCG CTTGCTGCAA
    TTGTCCTAAA AGCGGACGAC CCCGTTTGGT CGCACCTGGC GAACGACGTT

1001 CTCTCTCAGG GCCAGGCGGT GAAGGGCAAT CAGCTGTTGC CCGTCTCACT
    GAGAGAGTCC CGGTCCGCCA CTTCCCGTTA GTCGACAACG GGCAGAGTGA

1051 GGTGAAAAGA AAAACCACCC TGGCGCCCAA TACGCAAACC GCCTCTCCCC
    CCACTTTTCT TTTTGGTGGG ACCGCGGGTT ATGCGTTTGG CGGAGAGGGG

1101 GCGCGTTGGC CGATTCATTA ATGCAGCTGG CACGACAGGT TTCCCGACTG
    CGCGCAACCG GCTAAGTAAT TACGTCGACC GTGCTGTCCA AAGGGCTGAC

1151 GAAAGCGGGC AGTGAGCGGT ACCCGATAAA AGCGGCTTCC TGACAGGAGG
    CTTTCGCCCG TCACTCGCCA TGGGCTATTT TCGCCGAAGG ACTGTCTCTC

1201 CCGTTTTGTT TTGCAGCCCA CCTCAACGCA ATTAATGTGA GTTAGCTCAC
    GGCAAACAA AACGTCGGGT GGAGTTGCGT TAATTACACT CAATCGAGTG

1251 TCATTAGGCA CCCAGGCTT TACACTTTAT GCTTCCGGCT CGTATGTTGT
    AGTAATCCGT GGGGTCCGAA ATGTGAAATA CGAAGGCCGA GCATACAACA

1301 GTGGAATTGT GAGCGGATAA CAATTTTACA CAGGAAACAG CTATGACCAT
    CACCTTAACA CTCGCCTATT GTTAAAGTGT GTCCTTTGTC GATACTGGTA

          XbaI
          ~~~~~
1351 GATTACGAAT TTCTAGAGGT TGAGGTGATT TTATGAAAAA GAATATCGCA
    CTAATGCTTA AAGATCTCCA ACTCCACTAA AATACTTTTT CTTATAGCGT

1401 TTTCTTCTTG CATCTATGTT CGTTTTTCTT ATTGCTACAA ATGCATACGC
    AAAGAAGAAC GTAGATACAA GCAAAAAAGA TAACGATGTT TACGTATGCG

          EcoRI
          ~~~~~
1451 TGAATTCCAC CCAGAAACGC TGGTGAAAGT AAAAGATGCT GAAGATCAGT
    ACTTAAGGTG GGTCTTTGCG ACCACTTTCA TTTTCTACGA CTTCTAGTCA

1501 TGGGTGCACG AGTGGGTTAC ATCGAACTGG ATCTCAACAG CGGTAAGATC
    ACCCACGTGC TCACCCAATG TAGCTTGACC TAGAGTTGTC GCCATTCTAG

1551 CTTGAGAGTT TTCGCCCCGA AGAACGTTTT CCAATGATGA GCACTTTTAA
    GAACTCTCAA AAGCGGGGCT TCTTGCAAAA GGTTACTACT CGTGAAAATT

1601 AGTTCTGCTA TGTGGCGCGG TATTATCCCG TATTGACGCC GGGCAAGAGC
    TCAAGACGAT ACACCGCGCC ATAATAGGGC ATAAGTGGG CCCGTTCTCG

```

Figure 3: pBS vector series: functional map and sequence of pBS13 (continued)

|      |             |            |            |            |            |
|------|-------------|------------|------------|------------|------------|
|      |             |            |            |            | ScaI       |
|      |             |            |            |            | ~~~~~      |
| 1651 | AACTCGGTTCG | CCGCATACAC | TATTCTCAGA | ATGACTTGGT | TGAGTACTCA |
|      | TTGAGCCAGC  | GGCGTATGTG | ATAAGAGTCT | TACTGAACCA | ACTCATGAGT |
| 1701 | CCAGTCACAG  | AAAAGCATCT | TACGGATGGC | ATGACAGTAA | GAGAATTATG |
|      | GGTCAGTGTC  | TTTTCGTAGA | ATGCCTACCG | TACTGTCAAT | CTCTTAATAC |
| 1751 | CAGTGCTGCC  | ATAACCATGA | GTGATAACAC | TGCGGCCAAC | TTACTTCTGA |
|      | GTCACGACGG  | TATTGGTACT | CACTATTGTG | ACGCCGGTTG | AATGAAGACT |
| 1801 | CAACGATCGG  | AGGACCGAAG | GAGCTAACCG | CTTTTTTGCA | CAACATGGGG |
|      | GTTGCTAGCC  | TCCTGGCTTC | CTCGATTGGC | GAAAAAACGT | GTTGTACCCC |
| 1851 | GATCATGTAA  | CTCGCCTTGA | TCGTTGGGAA | CCGGAGCTGA | ATGAAGCCAT |
|      | CTAGTACATT  | GAGCGGAAGT | AGCAACCCCT | GGCCTCGACT | TACTTCGGTA |
| 1901 | ACCAAACGAC  | GAGCGTGACA | CCACGATGCC | TGTAGCAATG | GCAACAACGT |
|      | TGGTTTGCTG  | CTCGCACTGT | GGTGCTACGG | ACATCGTTAC | CGTTGTTGCA |
| 1951 | TGCGCAAAC   | ATTAAGTGGC | GAAGTACTTA | CTCTAGCTTC | CCGGCAACAA |
|      | ACGCGTTTGA  | TAATTGACCG | CTTGATGAAT | GAGATCGAAG | GGCCGTTGTT |
| 2001 | TTAATAGACT  | GGATGGAGGC | GGATAAAGTT | GCAGGACCAC | TTCTGCGCTC |
|      | AATTATCTGA  | CCTACCTCCG | CCTATTTCAA | CGTCCTGGTG | AAGACGCGAG |
| 2051 | GGCCCTTCCG  | GCTGGCTGGT | TTATTGCTGA | TAAATCTGGA | GCCGGTGAGC |
|      | CCGGGAAGGC  | CGACCGACCA | AATAACGACT | ATTTAGACCT | CGGCCACTCG |
| 2101 | GTGGGTCTCG  | CGGTATCATT | GCAGCACTGG | GGCCAGATGG | TAAGCCCTCC |
|      | CACCCAGAGC  | GCCATAGTAA | CGTCGTGACC | CCGGTCTACC | ATTCCGGAGG |
| 2151 | CGTATCGTAG  | TTATCTACAC | GACGGGGAGT | CAGGCAACTA | TGGATGAACG |
|      | GCATAGCATC  | AATAGATGTG | CTGCCCCTCA | GTCCGTTGAT | ACCTACTTGC |
| 2201 | AAATAGACAG  | ATCGCTGAGA | TAGGTGCCTC | ACTGATTAAG | CATTGGTAAT |
|      | TTTATCTGTC  | TAGCGACTCT | ATCCACGGAG | TGACTAATTC | GTAACCATTA |
|      |             | HindIII    |            |            |            |
|      |             | ~~~~~      |            |            |            |
| 2251 | GAGCATGCAA  | GCTTGACCTG | TGAAGTGAAA | AATGGCGCAC | ATTGTGCGAC |
|      | CTCGTACGTT  | CGAACTGGAC | ACTTCACTTT | TTACCGCGTG | TAACACGCTG |
| 2301 | ATTTTTTTTG  | TCTGCCGTTT | ACCGCTACTG | CGTCACGGAT | CCCCACGCGC |
|      | TAAAAAAAAC  | AGACGGCAAA | TGGCGATGAC | GCAGTGCCTA | GGGGTGCGCG |
| 2351 | CCTGTAGCGG  | CGCATTAAGC | GCGGCGGGTG | TGGTGGTTAC | GCGCAGCGTG |
|      | GGACATCGCC  | GCGTAATTCT | CGCCGCCAC  | ACCACCAATG | CGCGTCGCAC |
| 2401 | ACCGCTACAC  | TTGCCAGCGC | CCTAGCGCCC | GCTCCTTTCT | CTTTCTTCCC |
|      | TGGCGATGTG  | AACGGTCGCG | GGATCGCGGG | CGAGGAAAGC | GAAAGAAGGG |

Figure 3: pBS vector series: functional map and sequence of pBS13 (continued)

```

2451  TTCCTTTCTC GCCACGTTTC CCGGCTTTCC CCGTCAAGCT CTAAATCGGG
      AAGGAAAGAG CGGTGCAAGC GGCCGAAAGG GGCAGTTCGA GATTTAGCCC

2501  GCATCCCTTT AGGGTTCCGA TTTAGTGCTT TACGGCACCT CGACCCCAAA
      CGTAGGGAAA TCCCAAGGCT AAATCACGAA ATGCCGTGGA GCTGGGGTTT

2551  AAACCTTGATT AGGGTGATGG TTCACGTAGT GGGCCATCGC CCTGATAGAC
      TTTGAATAA TCCCACTACC AAGTGCATCA CCCGGTAGCG GGACTATCTG

2601  GGTTTTTCGC CCTTTGACGT TGGAGTCCAC GTTCTTTAAT AGTGGACTCT
      CCAAAAAGCG GGAAACTGCA ACCTCAGGTG CAAGAAATTA TCACCTGAGA

2651  TGTTCCAAAC TGGAACAACA CTCAACCCTA TCTCGGTCTA TTCTTTTGAT
      ACAAGGTTTG ACCTTGTTGT GAGTTGGGAT AGAGCCAGAT AAGAAAATA

2701  TTATAAGGGA TTTTGCCGAT TTCGGCCTAT TGGTTAAAAA ATGAGCTGAT
      AATATTCCCT AAAACGGCTA AAGCCGGATA ACCAATTTTT TACTCGACTA

2751  TTAACAAAAA TTTAACGCGA ATTTTAACAA AATATTAACG TTTACAATTT
      AATTGTTTTT AAATTGCGCT TAAAATTGTT TTATAATTGC AAATGTTAA

2801  CAGGTGGCAC TTTTCGGGGA AATGTGCGCG GAACCCCTAT TTGTTTATTT
      GTCCACCGTG AAAAGCCCCT TTACACGCGC CTTGGGGATA AACAAATAAA

2851  TTCTAAATAC ATTCAAATAT GTATCCGCTC ATGTCGAGAC GTTGGGTGAG
      AAGATTTATG TAAGTTTATA CATAGGCGAG TACAGCTCTG CAACCCACTC

2901  GTTCCAACCT TCACCATAAT GAAATAAGAT CACTACCGGG CGTATTTTTT
      CAAGGTTGAA AGTGGTATTA CTTTATTCTA GTGATGGCCC GCATAAAAAA

2951  GAGTTATCGA GATTTTCAGG AGCTAAGGAA GCTAAAATGG AGAAAAAAT
      CTCAATAGCT CTAAAAGTCC TCGATTCCCT CGATTTTACC TCTTTTTTTA

3001  CACTGGATAT ACCACCGTTG ATATATCCCA ATGGCATCGT AAAGAACATT
      GTGACCTATA TGGTGGCAAC TATATAGGGT TACCGTAGCA TTTCTTGTA

3051  TTGAGGCATT TCAGTCAGTT GCTCAATGTA CCTATAACCA GACCGTTCAG
      AACTCCGTAA AGTCAGTCAA CGAGTTACAT GGATATTGGT CTGGCAAGTC

3101  CTGGATATTA CGGCCTTTTT AAAGACCGTA AAGAAAAATA AGCACAAGTT
      GACCTATAAT GCCGGAAAAA TTTCTGGCAT TTCTTTTTAT TCGTGTTCAA

3151  TTATCCGGCC TTTATTCACT TTCTTGCCCG CCTGATGAAT GCTCATCCGG
      AATAGGCCGG AAATAAGTGT AAGAACGGGC GGACTACTTA CGAGTAGGCC

3201  AGTTCCGTAT GGCAATGAAA GACGGTGAGC TGGTGATATG GGATAGTGTT
      TCAAGGCATA CCGTTACTTT CTGCCACTCG ACCACTATAC CCTATCACA

3251  CACCCTTGTT ACACCGTTTT CCATGAGCAA ACTGAAACGT TTTTCATCGCT
      GTGGGAACAA TGTGGCAAAA GGTACTCGTT TGACTTTGCA AAAGTAGCGA

```

Figure 3: pBS vector series: functional map and sequence of pBS13 (continued)

|      |            |            |            |             |             |
|------|------------|------------|------------|-------------|-------------|
| 3301 | CTGGAGTGAA | TACCACGACG | ATTTCCGGCA | GTTTCTACAC  | ATATATTTCGC |
|      | GACCTCACTT | ATGGTGCTGC | TAAAGGCCGT | CAAAGATGTG  | TATATAAGCG  |
| 3351 | AAGATGTGGC | GTGTTACGGT | GAAAACCTGG | CCTATTTCCC  | TAAAGGGTTT  |
|      | TTCTACACCG | CACAATGCCA | CTTTTGGACC | GGATAAAGGG  | ATTTCCCAA   |
| 3401 | ATTGAGAATA | TGTTTTTCGT | CTCAGCCAAT | CCCTGGGTGA  | GTTTCACCAG  |
|      | TAACTCTTAT | ACAAAAAGCA | GAGTCGGTTA | GGGACCCACT  | CAAAGTGGTC  |
| 3451 | TTTTGATTTA | AACGTGGCCA | ATATGGACAA | CTTCTTCGCC  | CCCGTTTTCA  |
|      | AAACTAAAT  | TTGCACCGGT | TATACCTGTT | GAAGAAGCGG  | GGGCAAAAGT  |
| 3501 | CCATGGGCAA | ATATTATACG | CAAGGCGACA | AGGTGCTGAT  | GCCGCTGGCG  |
|      | GGTACCCGTT | TATAATATGC | GTTCCGCTGT | TCCACGACTA  | CGGCGACCGC  |
| 3551 | ATTCAGGTTT | ATCATGCCGT | CTGTGATGGC | TTCCATGTGC  | GCAGAATGCT  |
|      | TAAGTCCAAG | TAGTACGGCA | GACACTACCG | AAGGTACAGC  | CGTCTTACGA  |
|      |            | ScaI       |            |             |             |
|      |            | ~~~~~      |            |             |             |
| 3601 | TAATGAATTA | CAACAGTACT | GCGATGAGTG | GCAGGGCGGG  | GCGTAATTTT  |
|      | ATTACTTAAT | GTTGTCATGA | CGCTACTCAC | CGTCCC GCCC | CGCATTAATA  |
| 3651 | TTTAAGGCAG | TTATTGGTGC | CCTTAAACGC | CTGGTGCTAC  | GCCTGAATAA  |
|      | AAATTCCGTC | AATAACCACG | GGAATTTGCG | GACCACGATG  | CGGACTTATT  |
| 3701 | GTGATAATAA | GCGGATGAAT | GGCAGAAATT | CGAAAGCAAA  | TTCGACCCGG  |
|      | CACTATTATT | CGCCTACTTA | CCGTCTTTAA | GCTTTCGTTT  | AAGCTGGGCC  |
| 3751 | TCGTCGGTTC | AGGGCAGGGT | CGTTAAATAG | CCGCTTATGT  | CTATTGCTGG  |
|      | AGCAGCCAAG | TCCCGTCCCA | GCAATTTATC | GGCGAATACA  | GATAACGACC  |
| 3801 | TTTACCGGTT | TATTGACTAC | CGGAAGCAGT | GTGACCGTGT  | GCTTCTCAAA  |
|      | AAATGGCCAA | ATAACTGATG | GCCTTCGTCA | CACTGGCACA  | CGAAGAGTTT  |
| 3851 | TGCCTGAGGC | CAGTTTGCTC | AGGCTCTCCC | CGTGGAGGTA  | ATAATTGCTC  |
|      | ACGGACTCCG | GTCAAACGAG | TCCGAGAGGG | GCACCTCCAT  | TATTAACGAG  |
| 3901 | GACATGACCA | AAATCCCTTA | ACGTGAGTTT | TCGTTCCACT  | GAGCGTCAGA  |
|      | CTGTACTGGT | TTAGGGAAT  | TGCACTCAAA | AGCAAGGTGA  | CTCGCAGTCT  |
| 3951 | CCCCGTAGAA | AAGATCAAAG | GATCTTCTTG | AGATCCTTTT  | TTTCTGCGCG  |
|      | GGGGCATCTT | TTCTAGTTTC | CTAGAAGAAC | TCTAGGAAAA  | AAAGACGCGC  |
| 4001 | TAATCTGCTG | CTTGCAAACA | AAAAAACCAC | CGCTACCAGC  | GGTGGTTTGT  |
|      | ATTAGACGAC | GAACGTTTGT | TTTTTTGGTG | GCGATGGTCG  | CCACCAAACA  |
| 4051 | TTGCCGGATC | AAGAGCTACC | AACTCTTTTT | CCGAAGGTAA  | CTGGCTTCAG  |
|      | AACGGCCTAG | TTCTCGATGG | TTGAGAAAAA | GGCTTCCATT  | GACCGAAGTC  |

Figure 3: pBS vector series: functional map and sequence of pBS13 (continued)

```

4101  CAGAGCGCAG ATACCAAATA CTGTCCTTCT AGTGTAGCCG TAGTTAGGCC
      GTCTCGCGTC TATGGTTTAT GACAGGAAGA TCACATCGGC ATCAATCCGG

4151  ACCACTTCAA GAACTCTGTA GCACCGCCTA CATACTCGC TCTGCTAATC
      TGGTGAAGTT CTTGAGACAT CGTGGCGGAT GTATGGAGCG AGACGATTAG

4201  CTGTTACCAG TGGCTGCTGC CAGTGGCGAT AAGTCGTGTC TTACCGGGTT
      GACAATGGTC ACCGACGACG GTCACCGCTA TTCAGCACAG AATGGCCCAA

4251  GGA CTCAAGA CGATAGTTAC CGGATAAGGC GCAGCGGTCG GGCTGAACGG
      CCTGAGTTCT GCTATCAATG GCCTATTCCG CGTCGCCAGC CCGACTTGCC

4301  GGGGTTCGTG CACACAGCCC AGCTTGGAGC GAACGACCTA CACCGAACTG
      CCCC AAGCAC GTGTGTCGGG TCGAACCTCG CTTGCTGGAT GTGGCTTGAC

4351  AGATACCTAC AGCGTGAGCT ATGAGAAAGC GCCACGCTTC CCGAAGGGAG
      TCTATGGATG TCGCACTCGA TACTCTTTCG CGGTGCGAAG GGCTTCCCTC

4401  AAAGGCGGAC AGGTATCCGG TAAGCGGCAG GGTCGGAACA GGAGAGCGCA
      TTTCCGCCTG TCCATAGGCC ATTGCGCGTC CCAGCCTTGT CCTCTCGCGT

4451  CGAGGGAGCT TCCAGGGGGA AACGCCTGGT ATCTTTATAG TCCTGTCTGGG
      GCTCCCTCGA AGGTCCCCCT TTGCGGACCA TAGAAATATC AGGACAGCCC

4501  TTTCGCCACC TCTGACTTGA GCGTCGATTT TTGTGATGCT CGTCAGGGGG
      AAAGCGGTGG AGACTGAACT CGCAGCTAAA AACACTACGA GCAGTCCCCC

4551  GCGGAGCCTA TGGAAAAACG CCAGCAACGC GGCCTTTTTA CGGTTCTTGG
      CGCCTCGGAT ACCTTTTTGC GGTCGTTGCG CCGGAAAAAT GCCAAGGACC

4601  CCTTTTGCTG GCCTTTTGCT CACATG
      GGAAAACGAC CGGAAAACGA GTGTAC

```

Figure 4: Co-existence of phagemids: results of restriction digest

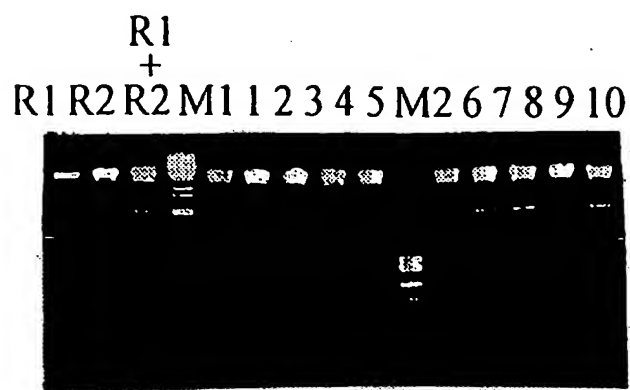




Figure 5: Phagemid vector pYING1-C1: functional map

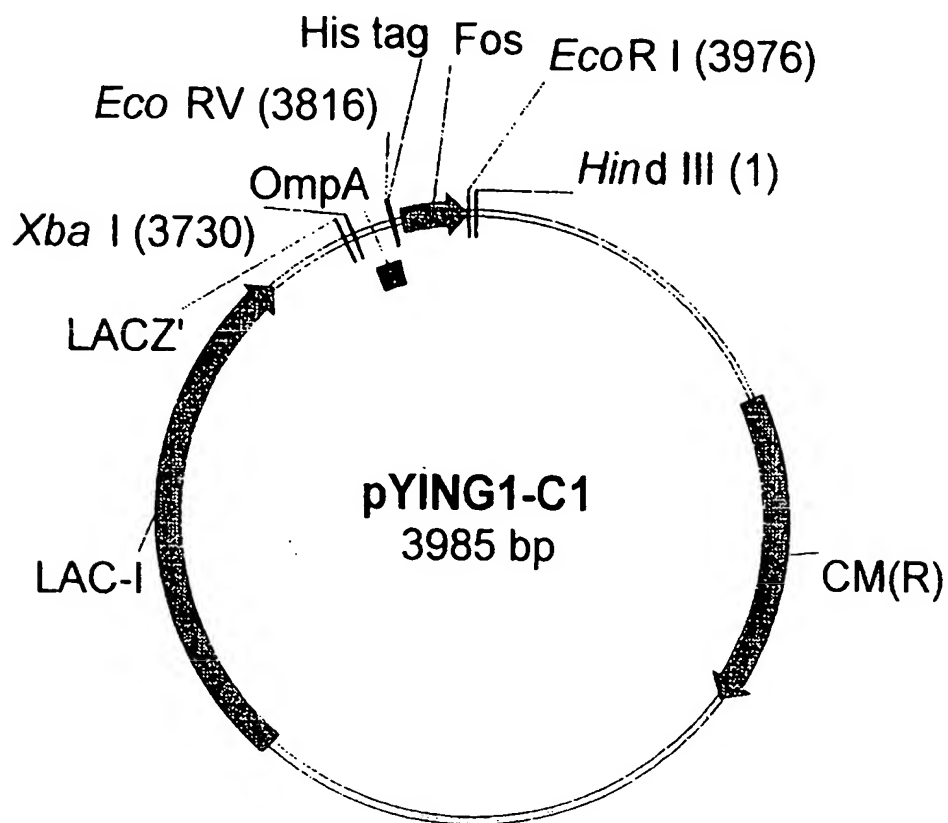


Figure 6: Phagemid vector pYANG3-A: functional map

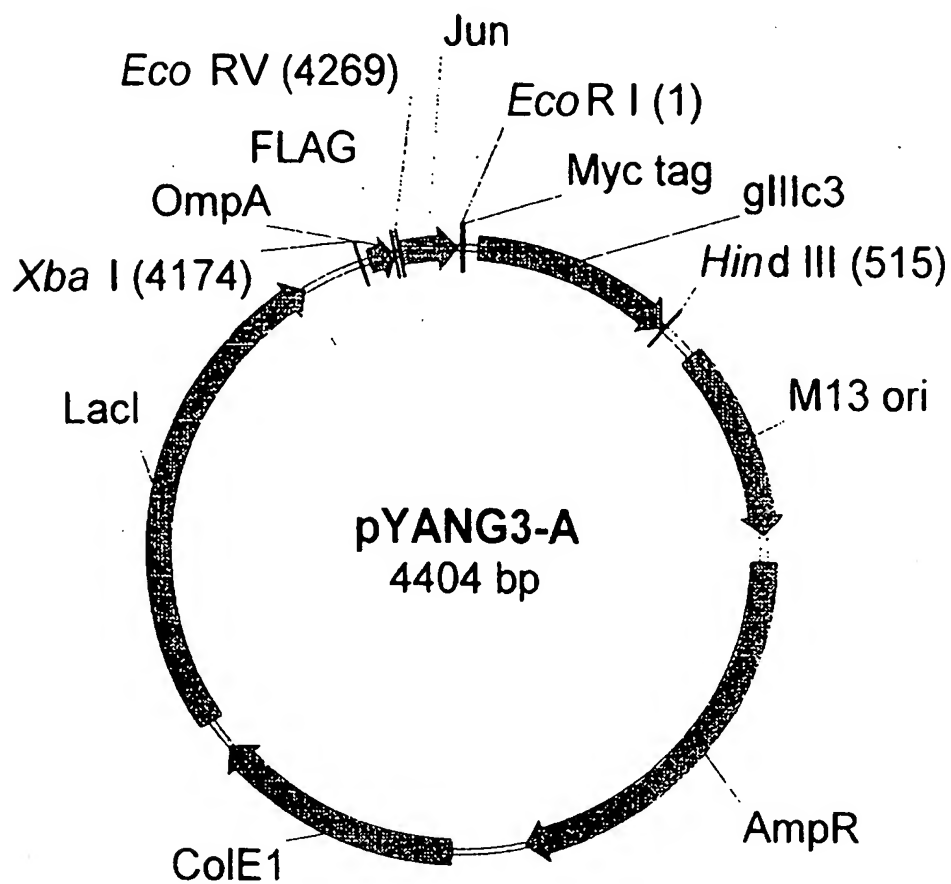
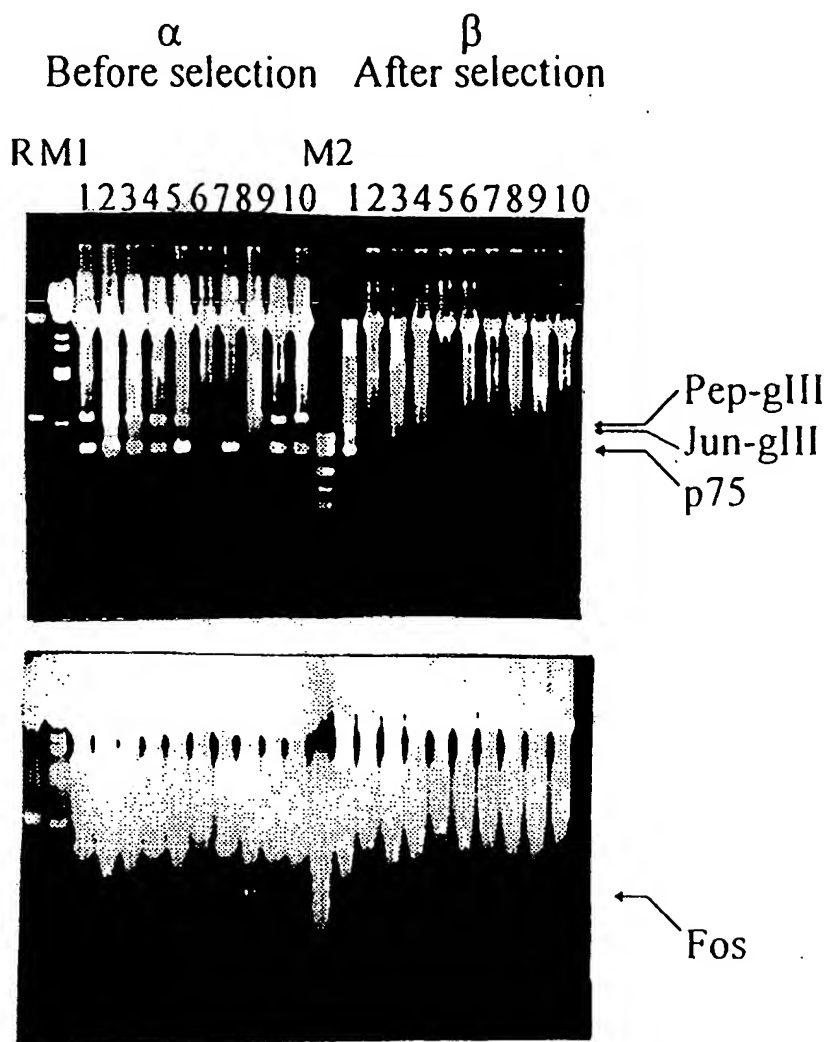


Figure 7: Analysis of selected clones (see Table 2)

7.a: Restriction digest of clones before and after selection



7.b: PCR of clones after selection with primers OPEP5L and OGIII3

$\beta$ : after selection

R1 R2 M1 2 3 4 5 6 7 8 9 10



Figure 8: Phagemid vector pING1-C1: functional map

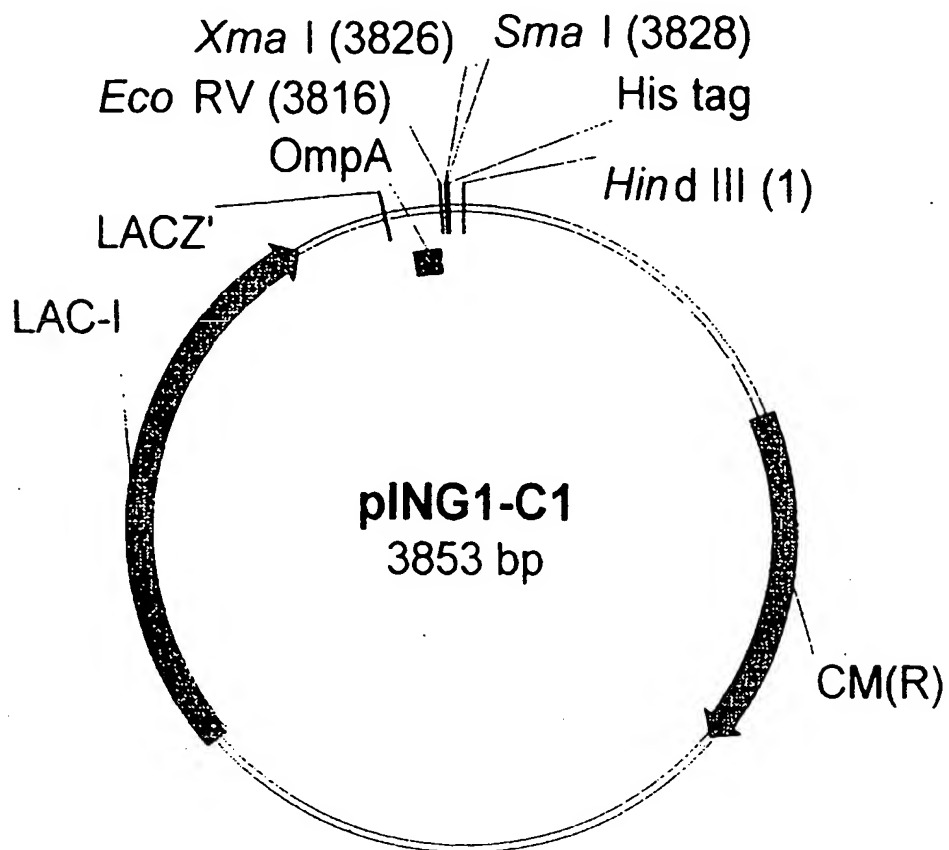


Figure 9: Phagemid vector pONG3-A: functional map

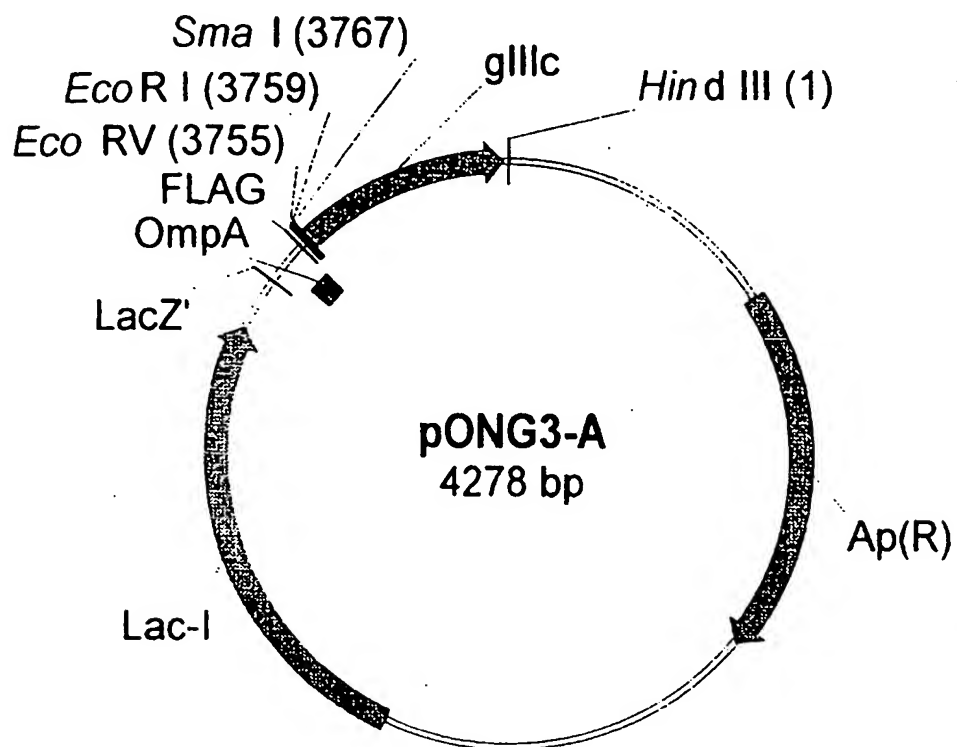


Figure 10: Co-transformation of phage and plasmid, polyphage formation and selection *via* SIP: general description

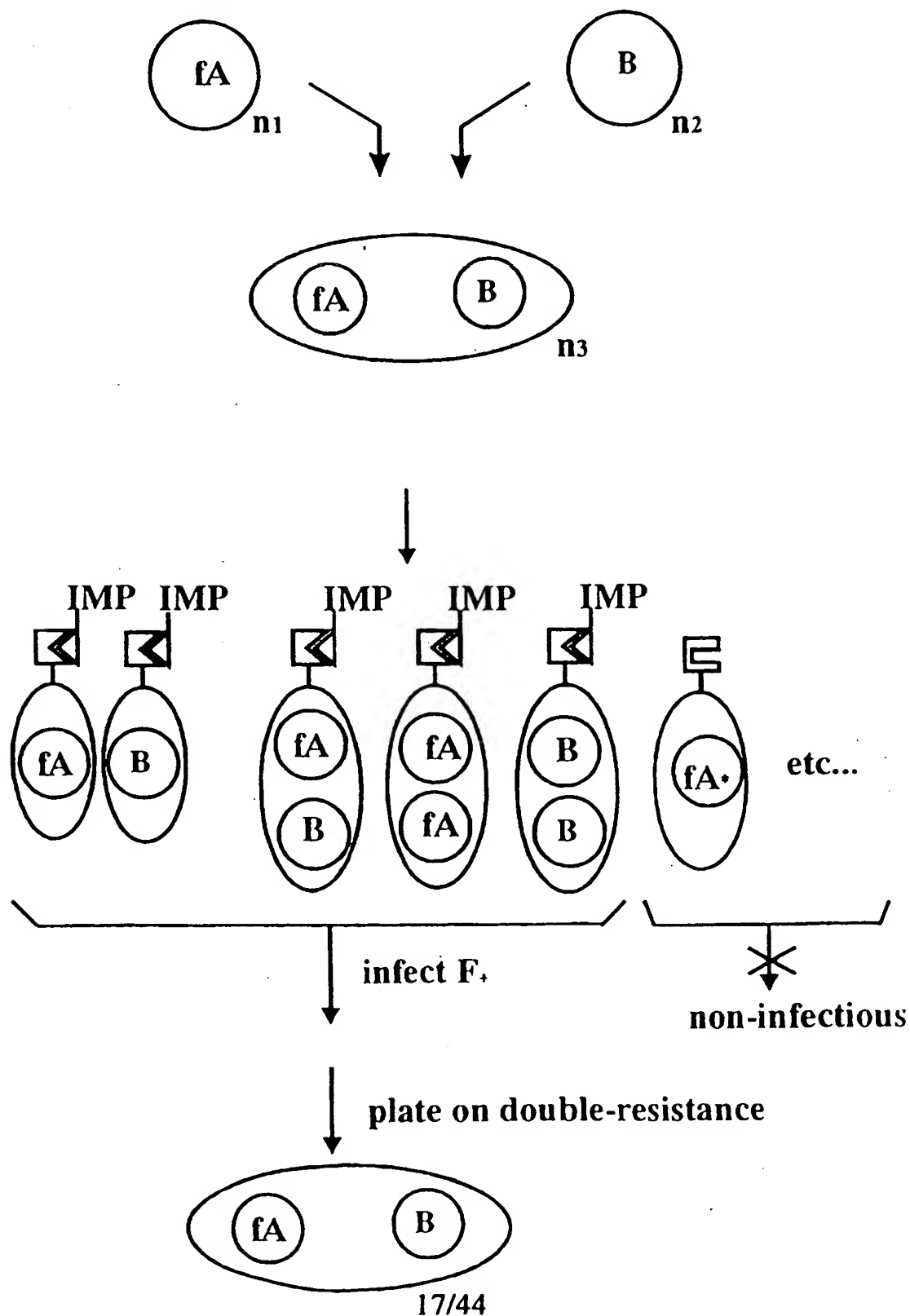


Figure 11: Phage vector fhag1A: functional map

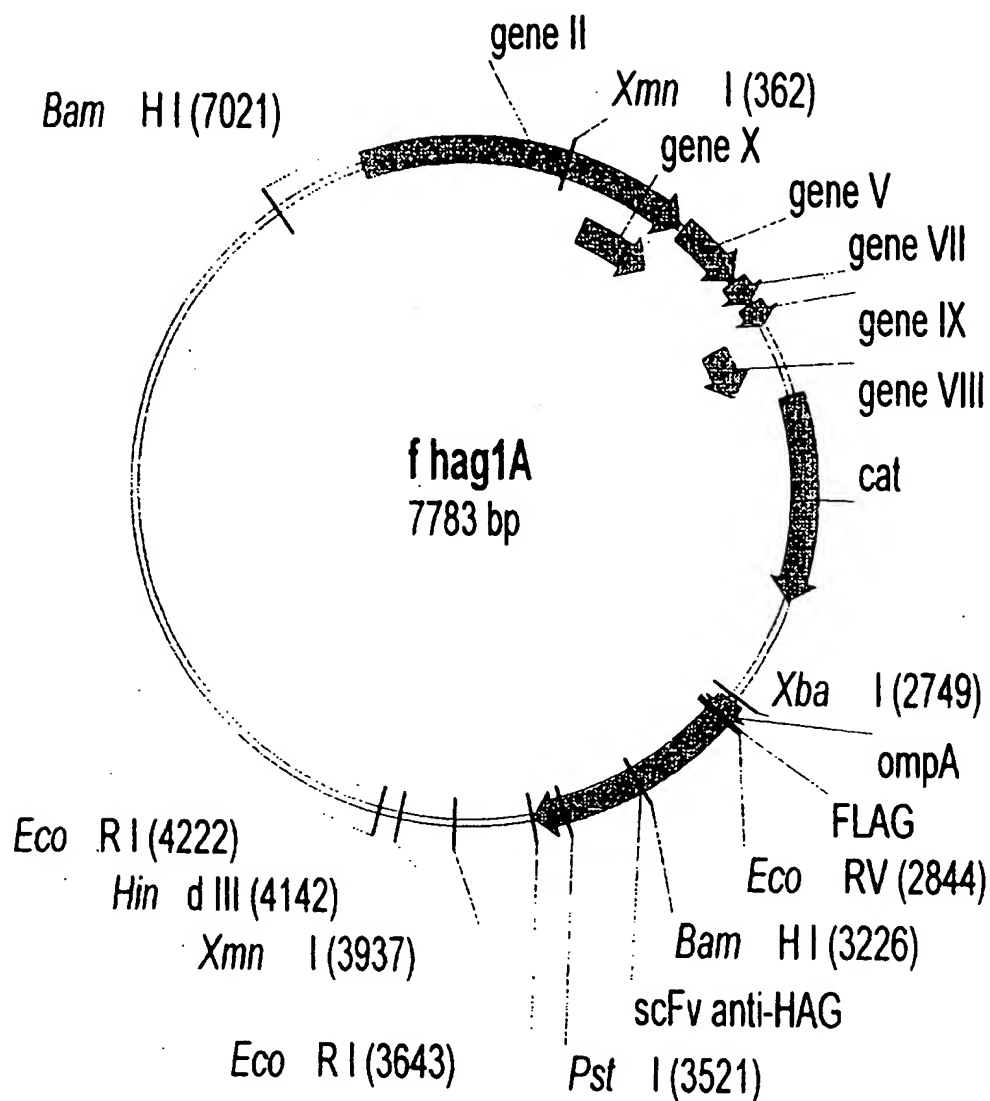


Figure 11a: CAT gene module: functional map and sequence

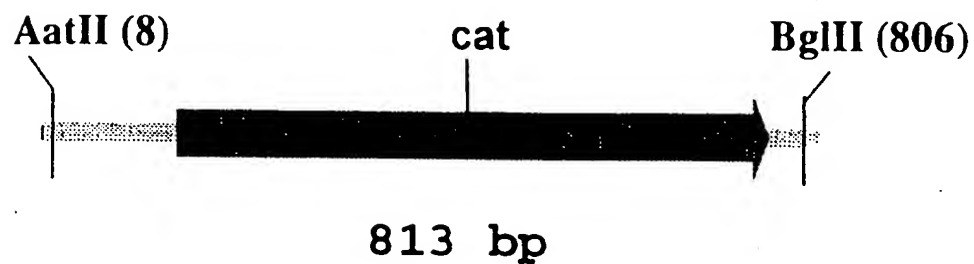




Figure 11a: CAT gene module: functional map and sequence  
(cont.)

|     |                           |                           |                           |                           |                           |
|-----|---------------------------|---------------------------|---------------------------|---------------------------|---------------------------|
|     | AatII<br>-----            |                           |                           |                           |                           |
| 1   | GGGACGTCGG<br>CCCTGCAGCC  | GTGAGGTTCC<br>CACTCCAAGG  | AACTTTCACC<br>TTGAAAGTGG  | ATAATGAAAT<br>TATTACTTTA  | AAGATCACTA<br>TTCTAGTGAT  |
| 51  | CCGGGCGTAT<br>GGCCCGCATA  | TTTTTGAGTT<br>AAAAACTCAA  | ATCGAGATTT<br>TAGCTCTAAA  | TCAGGAGCTA<br>AGTCCTCGAT  | AGGAAGCTAA<br>TCCTTCGATT  |
| 101 | AATGGAGAAA<br>TTACCTCTTT  | AAAATCACTG<br>TTTATAGTGAC | GATATACCAC<br>CTATATGGTG  | CGTTGATATA<br>GCAACTATAT  | TCCCAATGGC<br>AGGGTTACCG  |
| 151 | ATCGTAAAGA<br>TAGCATTCTT  | ACATTTTGAG<br>TGTA AAACTC | GCATTTTCAGT<br>CGTAAAGTCA | CAGTTGCTCA<br>GTCAACGAGT  | ATGTACCTAT<br>TACATGGATA  |
| 201 | AACCAGACCG<br>TTGGTCTGGC  | TTCAGCTGGA<br>AAGTCGACCT  | TATTACGGCC<br>ATAATGCCGG  | TTTTTAAAGA<br>AAAAATTTCT  | CCGTAAAGAA<br>GGCATTTCTT  |
| 251 | AAATAAGCAC<br>TTTATTCGTG  | AAGTTTTATC<br>TTCAAAAATAG | CGGCCTTTAT<br>GCCGGAAATA  | TCACATTCTT<br>AGTGTAAGAA  | GCCCCGCTGA<br>CGGGCGGACT  |
| 301 | TGAATGCTCA<br>ACTTACGAGT  | CCCGGAGTTC<br>GGGCCTCAAG  | CGTATGGCAA<br>GCATACCGTT  | TGAAAGACGG<br>ACTTTC TGCC | TGAGCTGGTG<br>ACTCGACCAC  |
| 351 | ATATGGGATA<br>TATACCCTAT  | GTGTTACACC<br>CACAAAGTGG  | TTGTTACACC<br>ACAATGTGG   | GTTTTCCATG<br>CAAAAGGTAC  | AGCAAACCTGA<br>TCGTTTGACT |
| 401 | AACGTTTTCA<br>TTGCAAAAAGT | TCGCTCTGGA<br>AGCGAGACCT  | GTGAATACCA<br>CACTTATGGT  | CGACGATTTT<br>GCTGCTAAAG  | CGGCAGTTTC<br>GCCGTCAAAG  |
| 451 | TACACATATA<br>ATGTGTATAT  | TTCGCAAGAT<br>AAGCGTCTTA  | GTGGCGTGTT<br>CACC GCACAA | ACGGTGAAAA<br>TGCCACTTTT  | CCTGGCCTAT<br>GGACCGGATA  |
| 501 | TTCCCTAAAG<br>AAGGGATTTT  | GGTTTATTTGA<br>CCAAATAACT | GAATATGTTT<br>CTTATACAAA  | TTGCTCTCAG<br>AAGCAGAGTC  | CCAATCCCTG<br>GGTTAGGGAC  |
| 551 | GGTGAGTTTC<br>CCTCTCAAAG  | ACCAGTTTTG<br>TGGTCAAAAC  | ATTTAAACGT<br>TAAATTTGCA  | AGCCAATATG<br>TCGGTTATAC  | GACAACTTCT<br>CTGTTGAAGA  |
| 601 | TCGCCCCCGT<br>AGCGGGGGCA  | TTTCACTATG<br>AAAGTGATAC  | GGCAAATATT<br>CCGTTTATAA  | ATACGCAAGG<br>TATGCGTTCC  | CGACAAGGTG<br>GCTGTTCCAC  |
| 651 | CTGATGCCGC<br>GACTACGGCG  | TGGCGATTCA<br>ACCGCTAAGT  | GGTTCATCAT<br>CCAAGTAGTA  | GCCGTTTG TG<br>CGGCAAACAC | ATGGCTTCCA<br>TACCGAAGGT  |
| 701 | TGTCGGCAGA<br>ACAGCCGTCT  | ATGCTTAATG<br>TACGAATTAC  | AATTACAACA<br>TTAATGTTGT  | GTA CTGCGAT<br>CATGACGCTA | GAGTGGCAGG<br>CTCACC GTCC |
| 751 | GCGGGGCGTA<br>CGCCCCGCAT  | ATTTTTTTTAA<br>TAAAAAAATT | GGCAGTTATT<br>CCGTCAATAA  | GGGTGCCCTT<br>CCCACGGGAA  | AAACGCCTGG<br>TTTGCGGACC  |
|     | BglII<br>-----            |                           |                           |                           |                           |
| 801 | TGCTAGATCT<br>ACGATCTAGA  | TCC<br>AGG                |                           |                           |                           |

Figure 12: Phage vector fjun1A: functional map

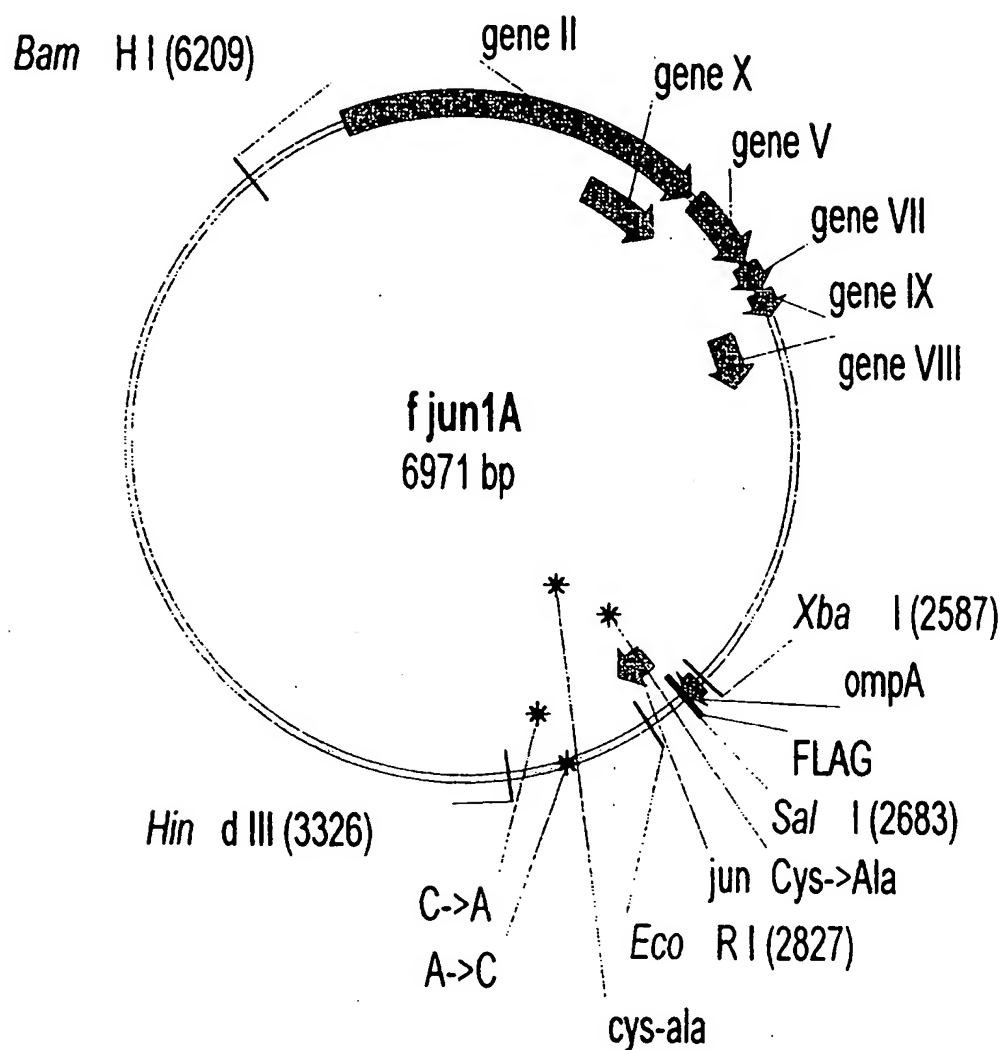


Figure 13: Phage vector fjun1B: functional map

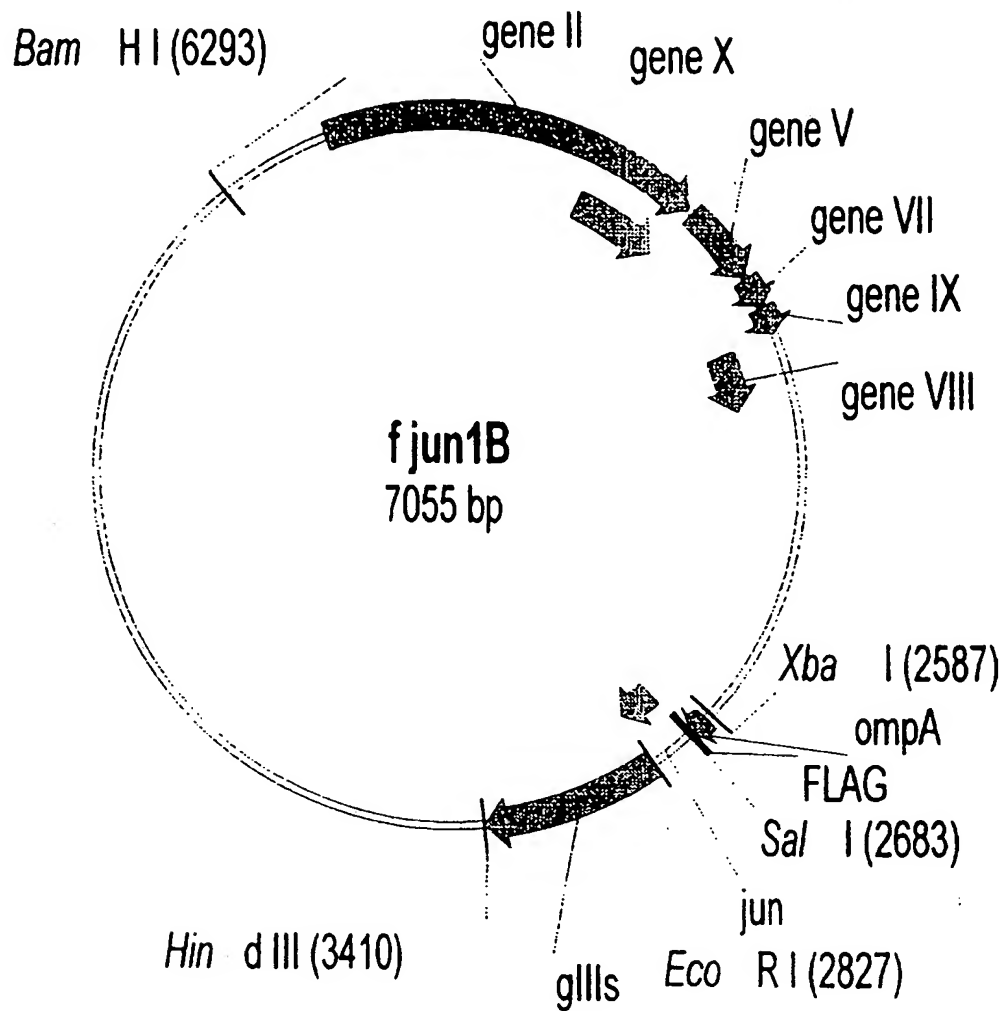


Figure 14: Phage vector fpep3\_1B: functional map

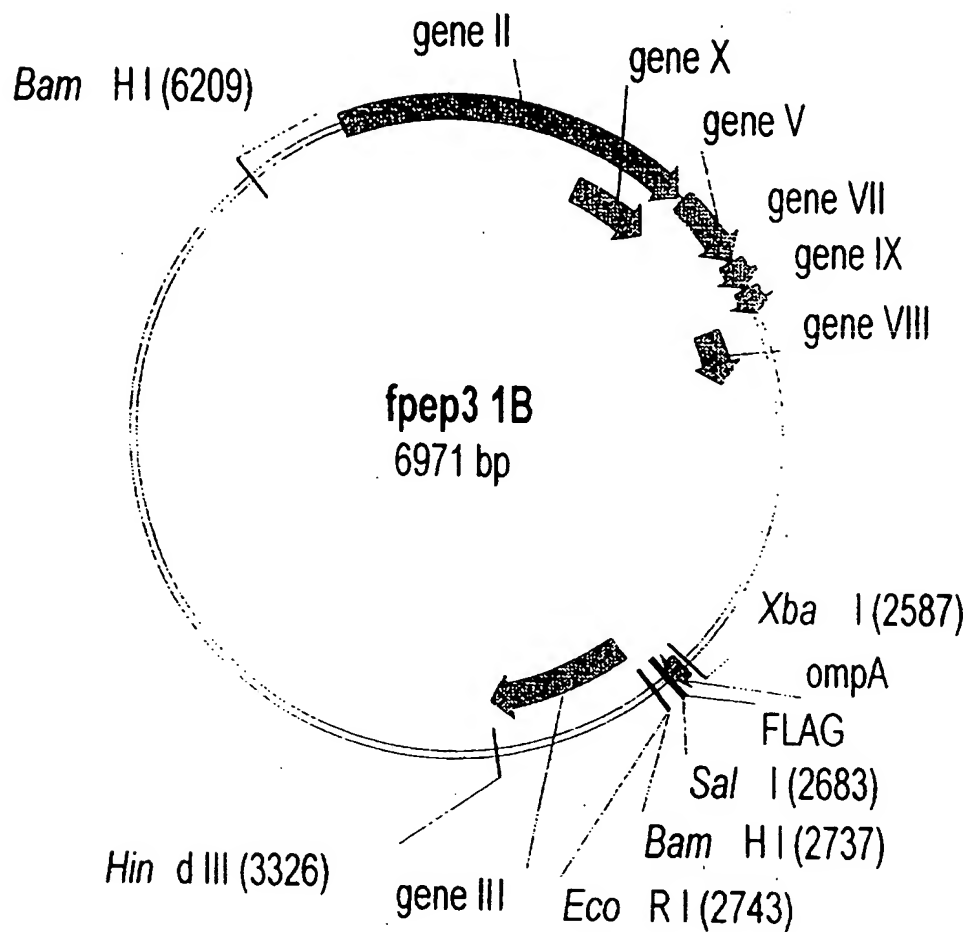


Figure 15: Phage vector fNGF\_1B: functional map

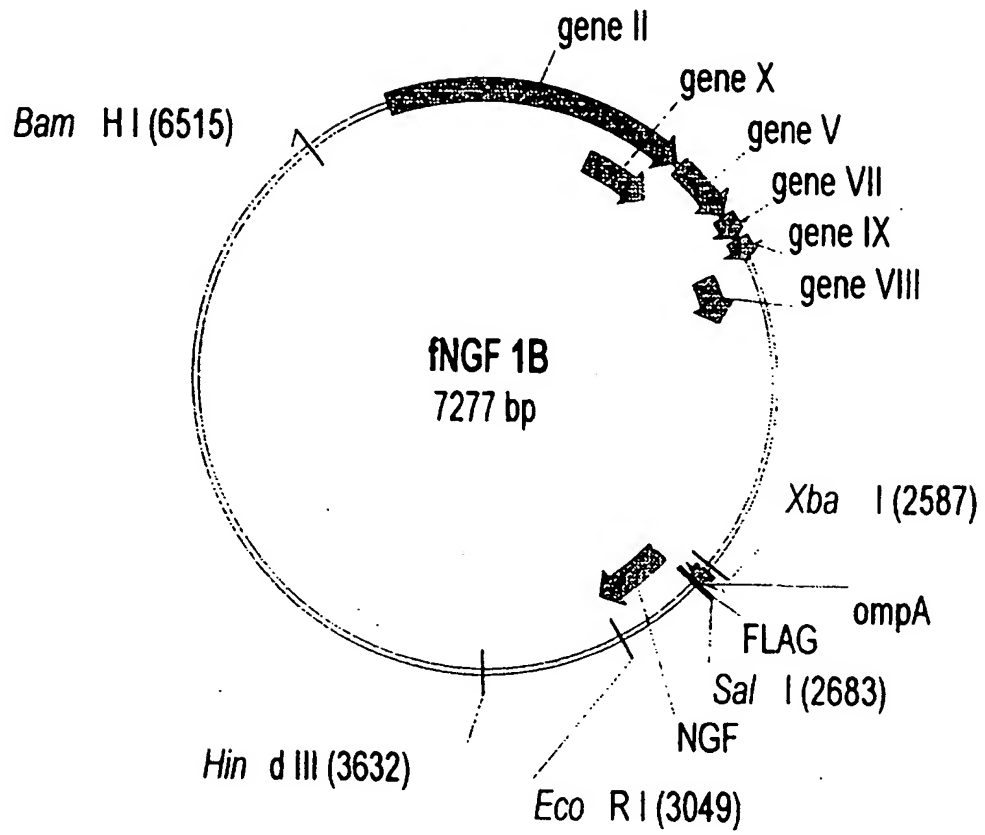


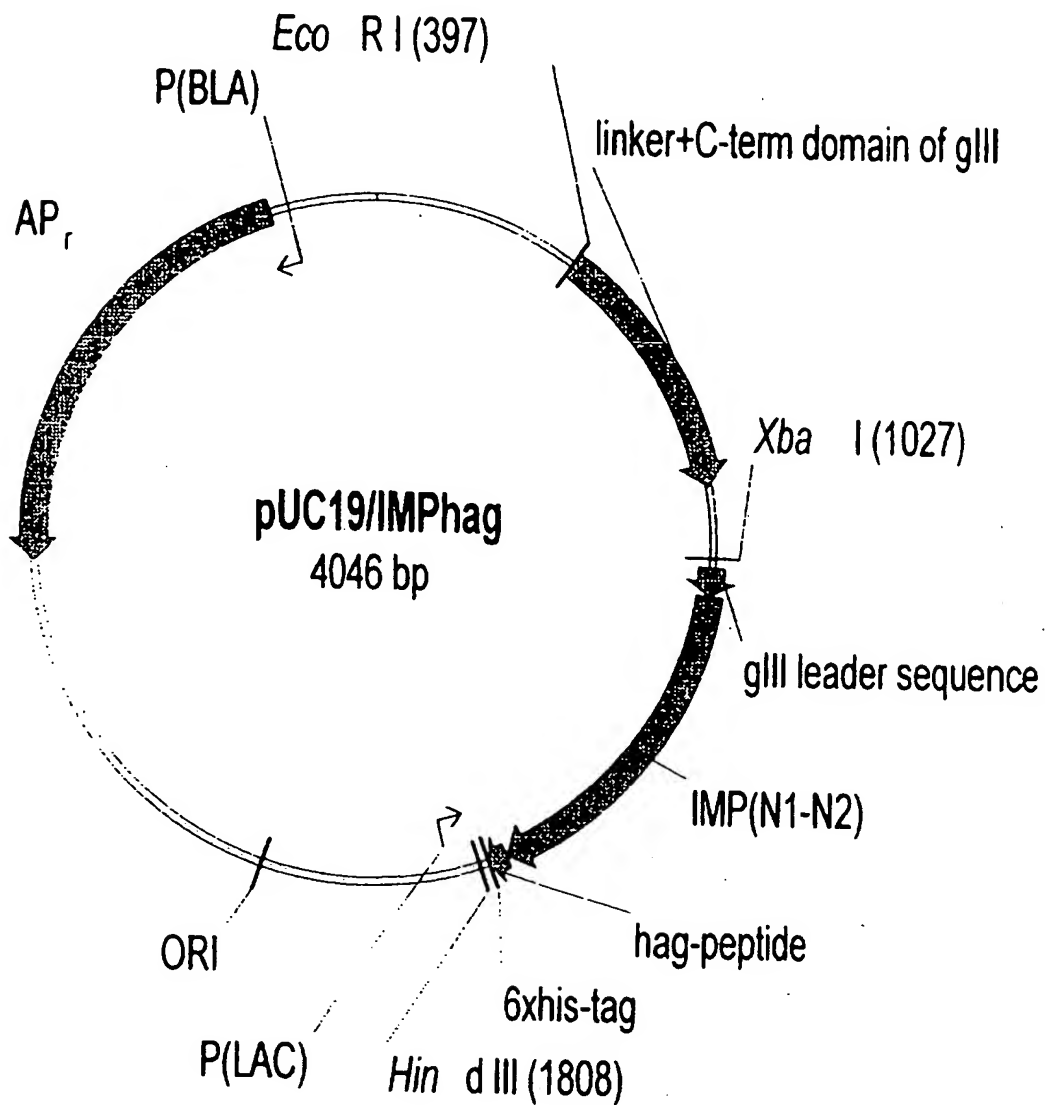
Figure 16: Plasmid pUC19/IMP<sub>hag</sub>: functional map

Figure 17: Plasmid pUC18/IMPp75: functional map

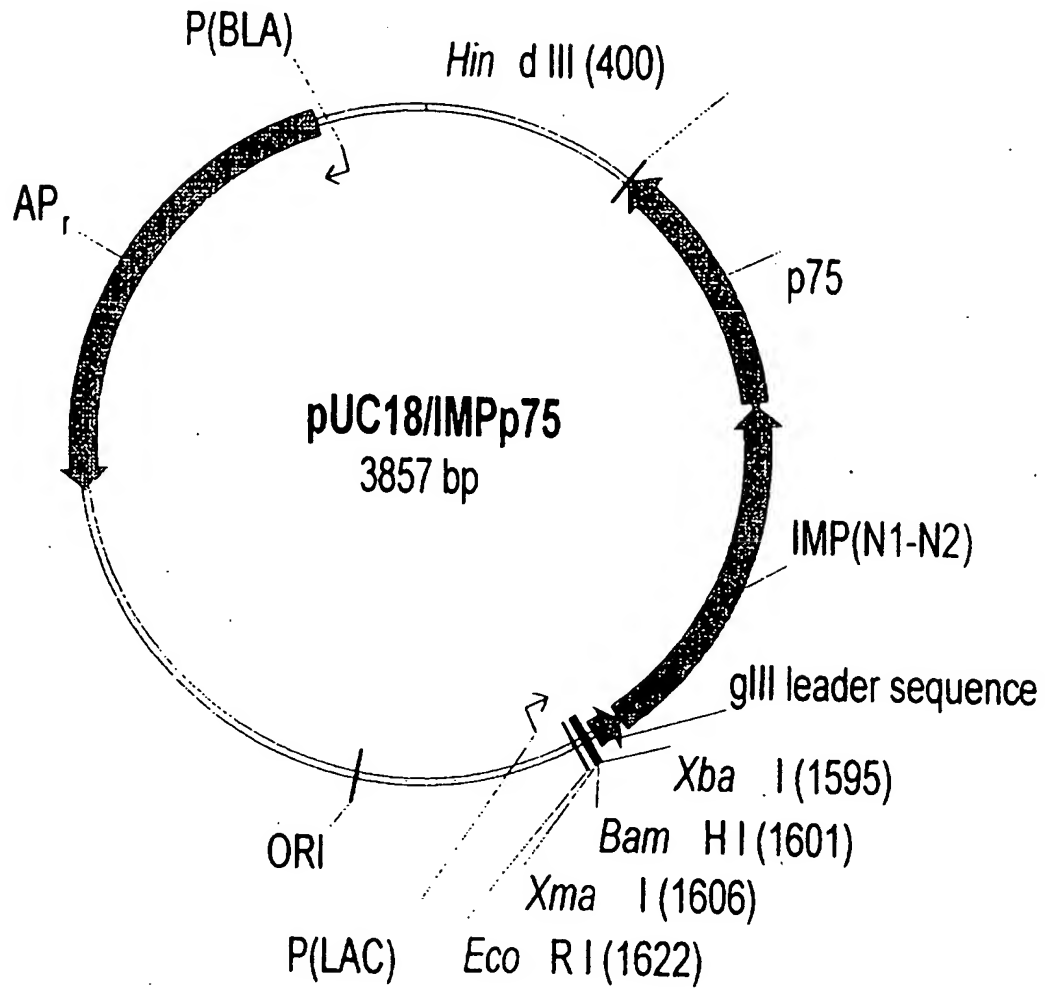


Figure 18: Plasmid pUC18/IMPIL16: functional map

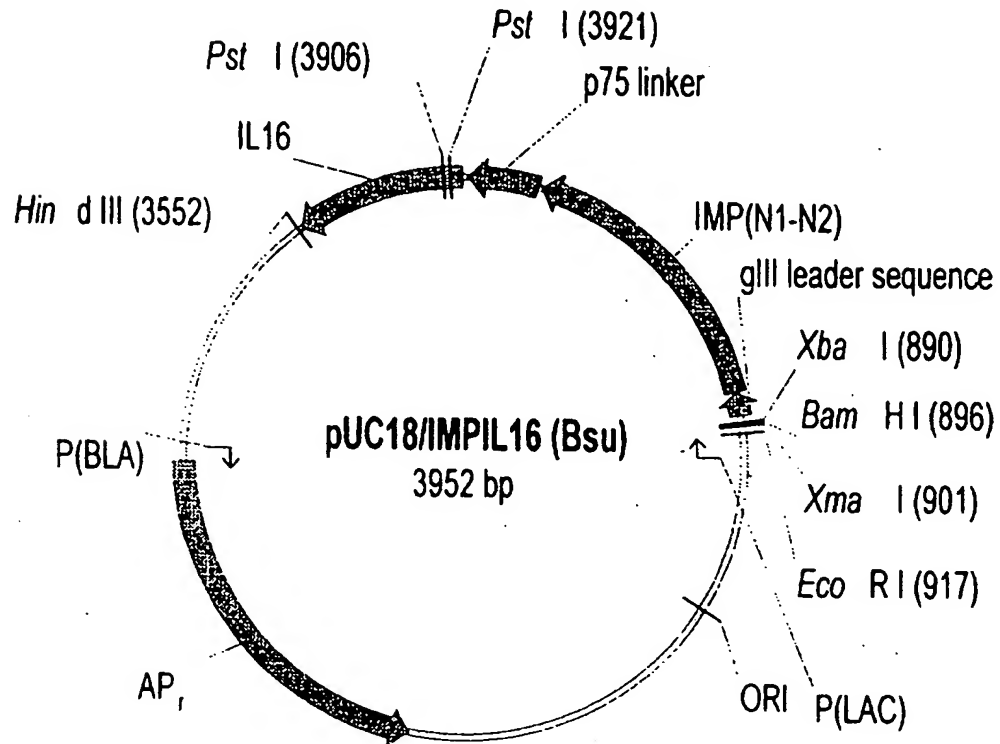




Figure 19: Analysis of selected clones (see Table 3)

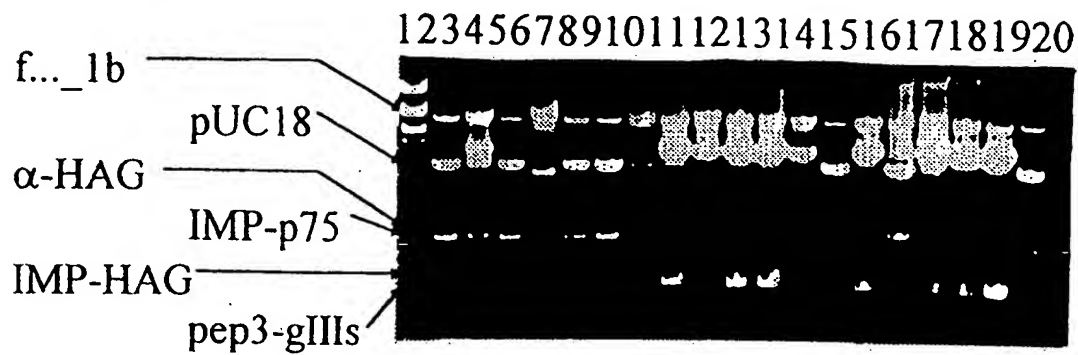


Figure 20: Co-transformation of phagemids, *in vivo* recombination and selection *via* His-tag: general description

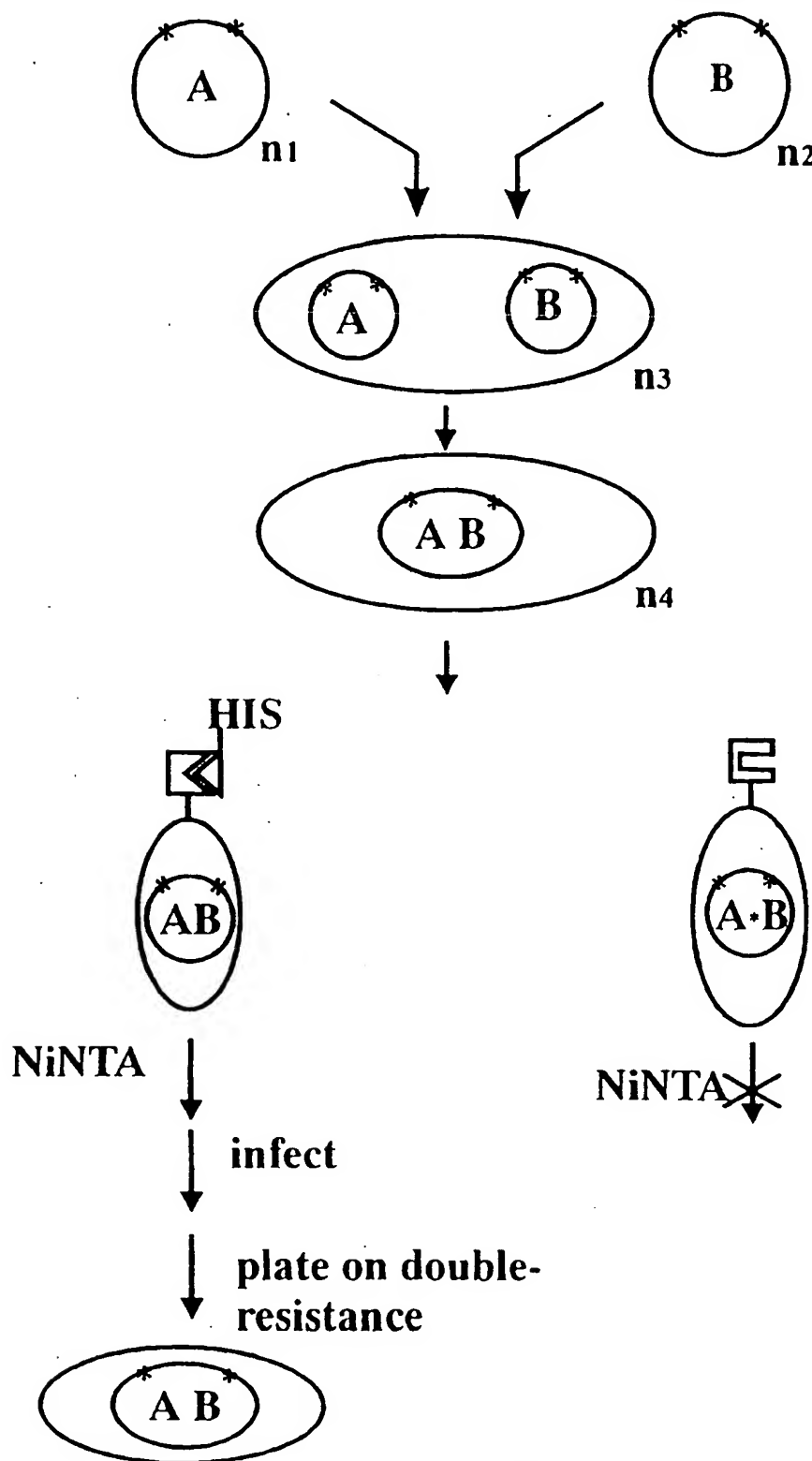


Figure 21: *In vitro* recombination and selection via His-tag: general description

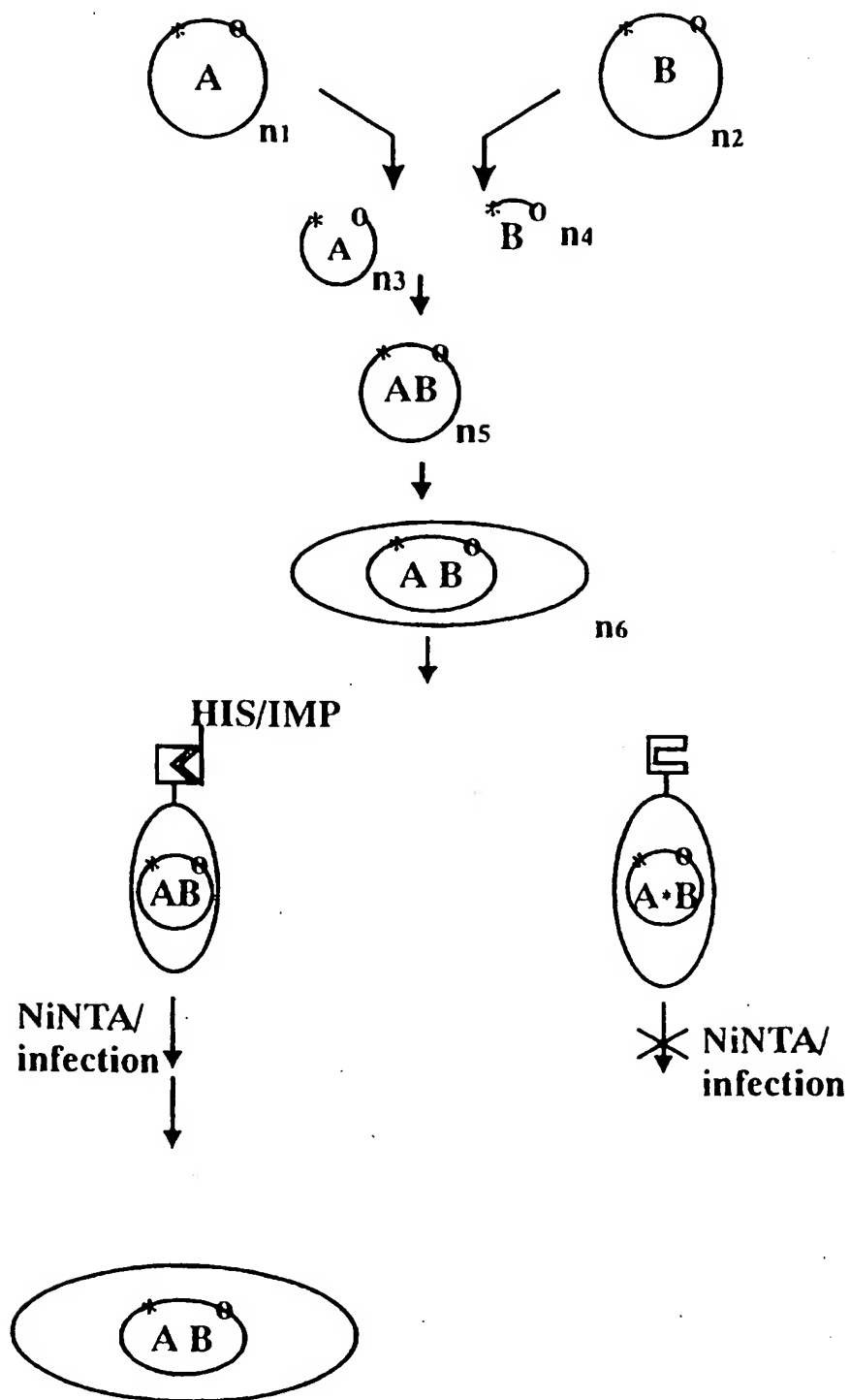


Figure 22: Phage vector fjunhag: functional map

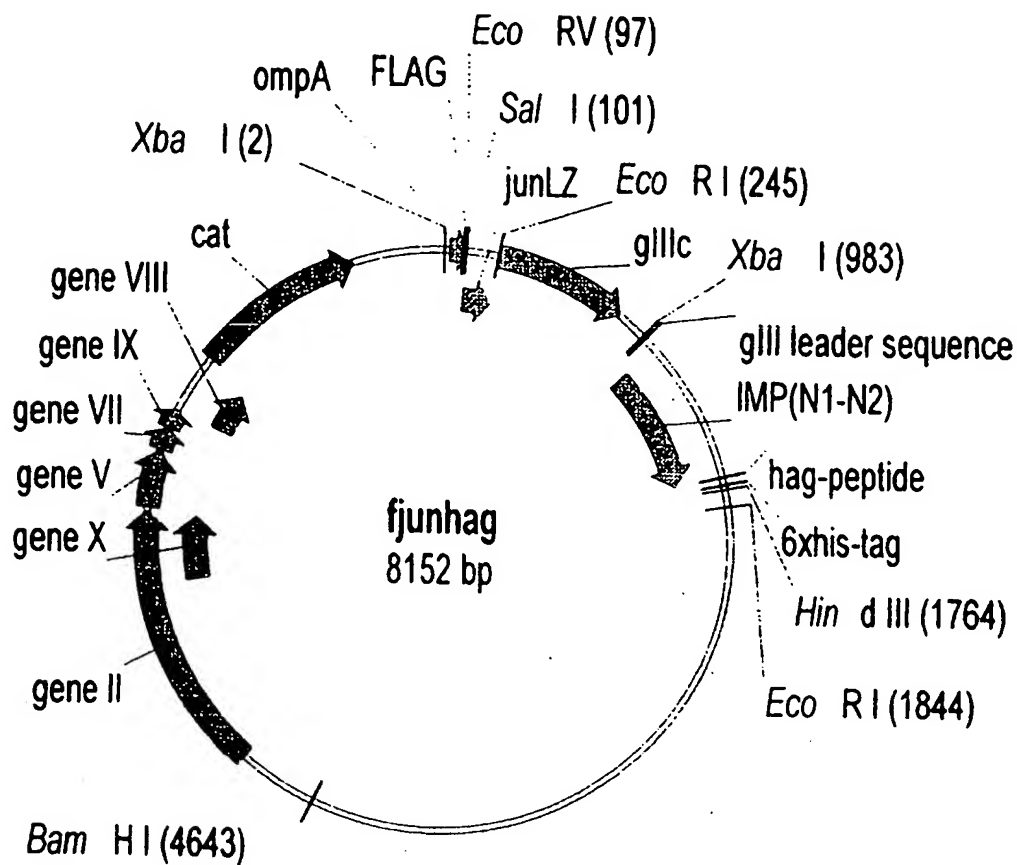


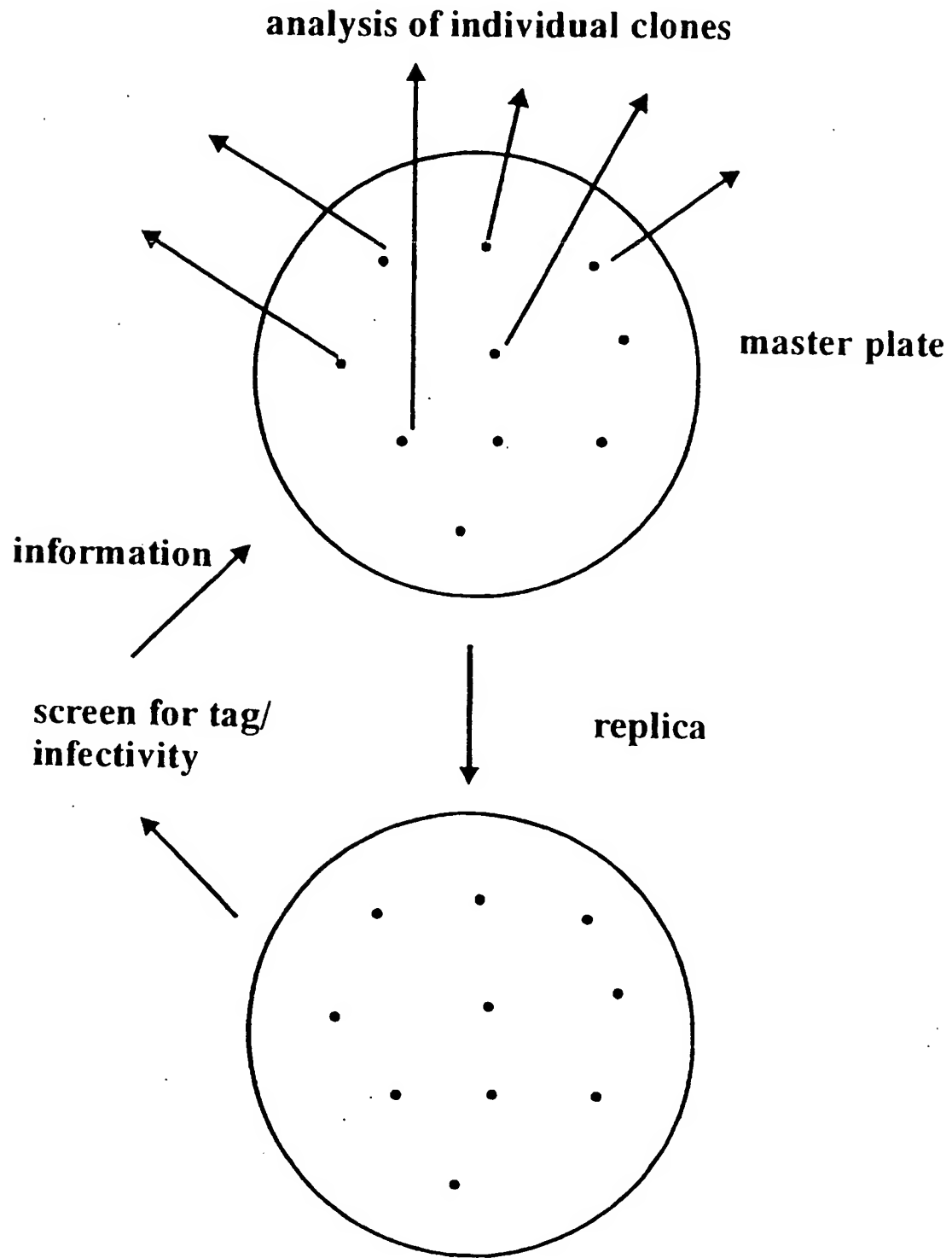
Figure 23: Spatial *in vivo* SIP: general description

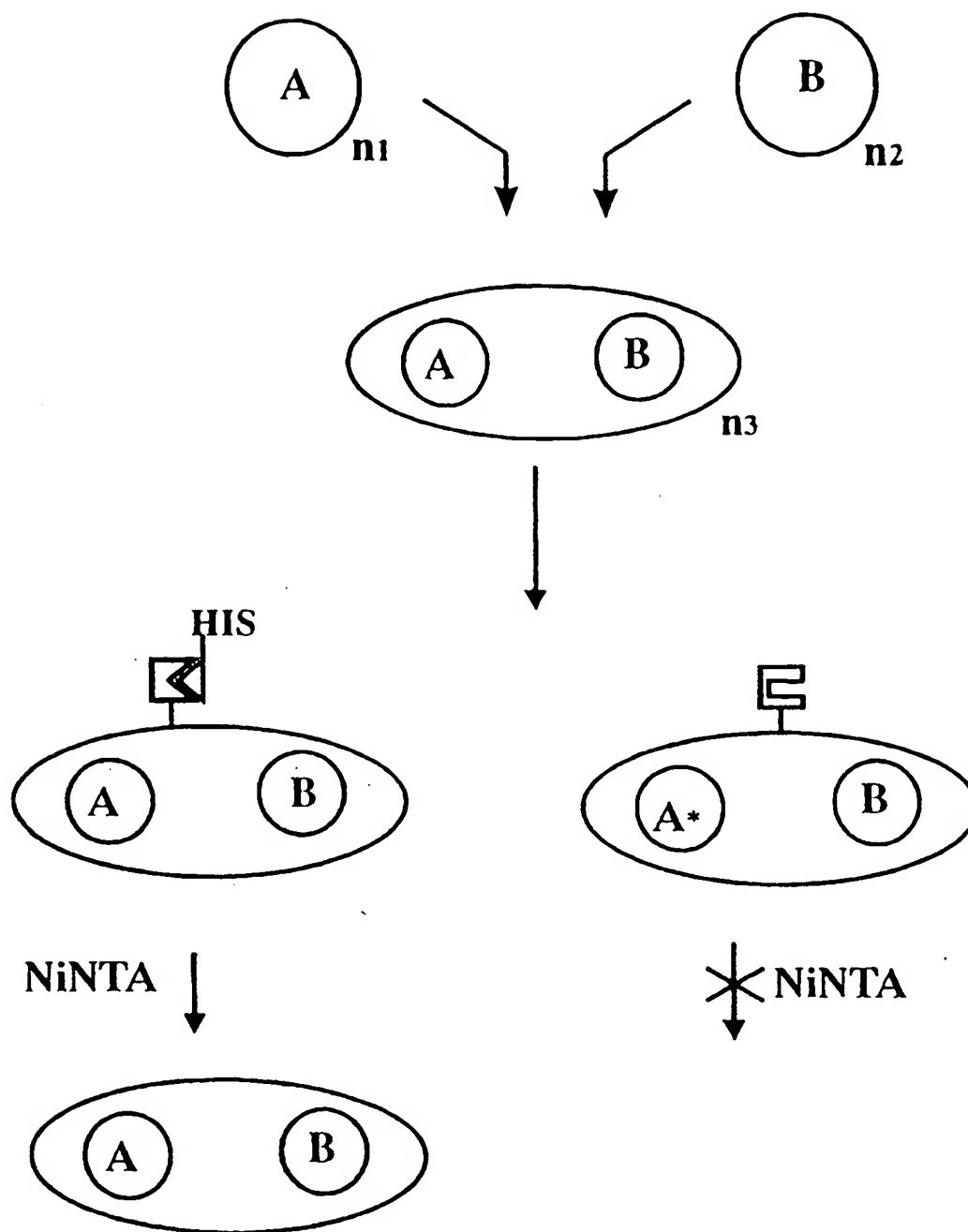
Figure 24: *E. coli* display: general description

Figure 25: pTERMsc2H10myc3sCAM: functional map and sequence

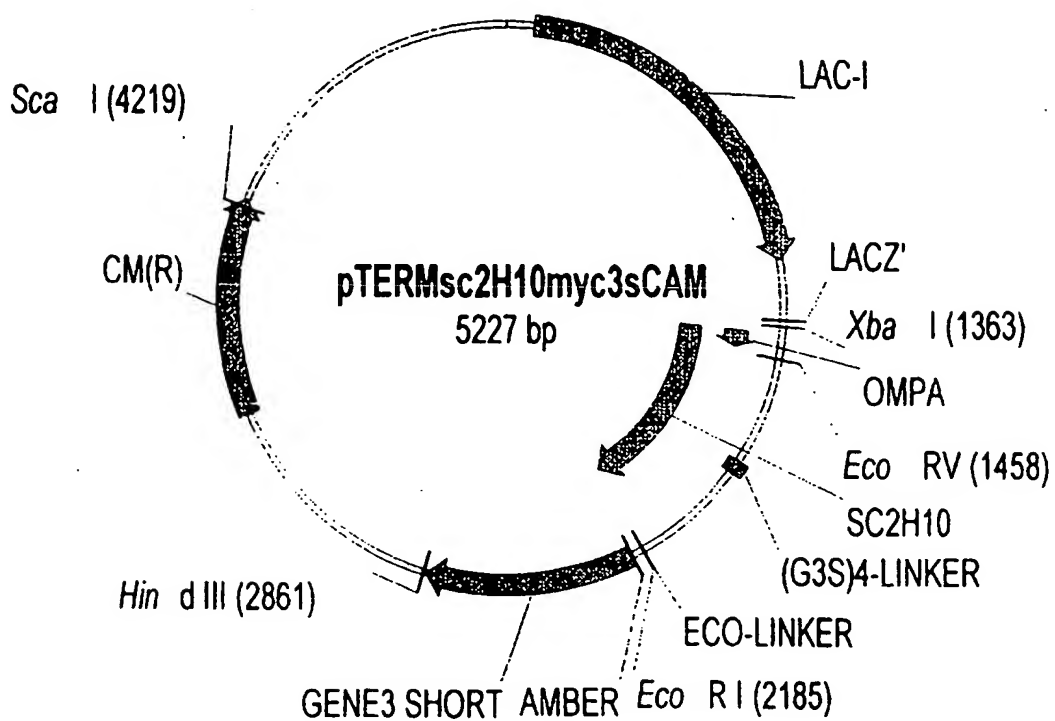


Figure 25: pTERMsc2H10myc3sCAM: functional map and sequence (continued)

```

1  ACCCGACACC ATCGAATGGC GCAAAACCTT TCGCGGTATG GCATGATAGC
   TGGGCTGTGG TAGCTTACCG CGTTTTGGAA AGCGCCATAC CGTACTATCG

51  GCCCGGAAGA GAGTCAATTC AGGGTGGTGA ATGTGAAACC AGTAACGTTA
   CGGGCCTTCT CTCAGTTAAG TCCCACCACT TACACTTTGG TCATTGCAAT

101 TACGATGTCG CAGAGTATGC CGGTGTCTCT TATCAGACCG TTTCCCGCGT
   ATGCTACAGC GTCTCATACG GCCACAGAGA ATAGTCTGGC AAAGGGCGCA

151 GGTGAACCAG GCCAGCCACG TTTCTGCGAA AACGCGGGAA AAAGTGAAG
   CCACCTGGTC CGGTCGGTGC AAAGACGCTT TTGCGCCCTT TTTACCTTC

201 CGGCGATGGC GGAGCTGAAT TACATTCCCA ACCGCGTGGC ACAACAACTG
   GCCGCTACCG CCTCGACTTA ATGTAAGGGT TGGCGCACCG TGTGTTGAC

251 GCGGGCAAAC AGTCGTTGCT GATTGGCGTT GCCACCTCCA GTCTGGCCCT
   CGCCCGTTTG TCAGCAACGA CTAACCGCAA CGGTGGAGGT CAGACCGGGA

301 GCACGCGCCG TCGCAAATTG TCGCGGCGAT TAAATCTCGC GCCGATCAAC
   CGTGCGCGGC AGCGTTTAAC AGCGCCGCTA ATTTAGAGCG CGGCTAGTTG

351 TGGGTGCCAG CGTGGTGGTG TCGATGGTAG AACGAAGCGG CGTCGAAGCC
   ACCCACGGTC GCACCACCAC AGTACCATC TTGCTTCGCC GCAGCTTCGG

401 TGTAAGCGG CGGTGCACAA TCTTCTCGCG CAACGCGTCA GTGGGCTGAT
   ACATTTCGCC GCCACGTGTT AGAAGAGCGC GTTGCGCAGT CACCCGACTA

451 CATTAACTAT CCGCTGGATG ACCAGGATGC CATTGCTGTG GAAGCTGCCT
   GTAATTGATA GGCGACCTAC TGGTCCTACG GTAACGACAC CTTCGACGGA

501 GCACTAATGT TCCGGCGTTA TTTCTTGATG TCTCTGACCA GACACCCATC
   CGTGATTACA AGGCCGCAAT AAAGAACTAC AGAGACTGGT CTGTGGGTAG

551 AACAGTATTA TTTTCTCCCA TGAAGACGGT ACGCGACTGG GCGTGGAGCA
   TTGTCATAAT AAAAGAGGGT ACTTCTGCCA TGCCTGACC CGCACCTCGT

601 TCTGGTCGCA TTGGGTCACC AGCAAATCGC GCTGTTAGCG GGCCATTAA
   AGACCAGCGT AACCAGTGG TCGTTTAGCG CGACAATCGC CCGGGTAATT

651 GTTCTGTCTC GGCGCGTCTG CGTCTGGCTG GCTGGCATAA ATATCTCACT
   CAAGACAGAG CCGCGCAGAC GCAGACCGAC CGACCGTATT TATAGAGTGA

701 CGCAATCAAA TTCAGCCGAT AGCGGAACGG GAAGGCGACT GGAGTGCCAT
   GCGTTAGTTT AAGTCGGCTA TCGCCTTGCC CTTCCGCTGA CCTCACGGTA

751 GTCCGGTTTT CAACAAACCA TGCAAATGCT GAATGAGGGC ATCGTTCCCA
   CAGGCCAAAA GTTGTTTGGT ACGTTTACGA CTTACTCCCG TAGCAAGGGT

801 CTGCGATGCT GGTTGCCAAC GATCAGATGG CGCTGGGCGC AATGCGCGCC
   GACGCTACGA CCAACGGTTG CTAGTCTACC GCGACCCGCG TTACGCGCGG

```



Figure 25: pTERMsc2H10myc3sCAM: functional map and sequence (continued)

```

851  ATTACCGAGT CCGGGCTGCG CGTTGGTGCG GACATCTCGG TAGTGGGATA
    TAATGGCTCA GGCCCGACGC GCAACCACGC CTGTAGAGCC ATCACCCAT

901  CGACGATACC GAAGACAGCT CATGTTATAT CCCGCCGTTA ACCACCATCA
    GCTGCTATGG CTTCTGTCGA GTACAATATA GGGCGGCAAT TGGTGGTAGT

951  AACAGGATTT TCGCCTGCTG GGGCAAACCA GCGTGGACCG CTTGCTGCAA
    TTGTCCTAAA AGCGGACGAC CCCGTTTGGT CGCACCTGGC GAACGACGTT

1001 CTCTCTCAGG GCCAGGCGGT GAAGGGCAAT CAGCTGTTGC CCGTCTCACT
    GAGAGAGTCC CGGTCCGCCA CTTCCCCGTTA GTCGACAACG GGCAGAGTGA

1051 GGTGAAAAGA AAAACCACCC TGGCGCCCAA TACGCAAACC GCCTCTCCCC
    CCACTTTTCT TTTTGGTGGG ACCGCGGGTT ATGCGTTTGG CGGAGAGGGG

1101 GCGCGTTGGC CGATTCAATTA ATGCAGCTGG CACGACAGGT TTCCCGACTG
    CGCGCAACCG GCTAAGTAAT TACGTCGACC GTGCTGTCCA AAGGGCTGAC

1151 GAAAGCGGGC AGTGAGCGGT ACCCGATAAA AGCGGCTTCC TGACAGGAGG
    CTTTCGCCCC TCACTCGCCA TGGGCTATTT TCGCCGAAGG ACTGTCCTCC

1201 CCGTTTTGTT TTGCAGCCCA CCTCAACGCA ATTAATGTGA GTTAGCTCAC
    GGCAAAACAA AACGTCGGGT GGAGTTGCGT TAATTACACT CAATCGAGTG

1251 TCATTAGGCA CCCCAGGCTT TACACTTTAT GCTTCCGGCT CGTATGTTGT
    AGTAATCCGT GGGGTCCGAA ATGTGAAATA CGAAGGCCGA GCATACAACA

1301 GTGGAATTGT GAGCGGATAA CAATTTACCA CAGGAAACAG CTATGACCAT
    CACCTTAACA CTCGCCTATT GTTAAAGTGT GTCCTTTGTC GATACTGGTA

                                XbaI
                                ~~~~~
1351 GATTACGAAT TTCTAGATAA CGAGGGCAAA AAATGAAAAA GACAGCTATC
    CTAATGCTTA AAGATCTATT GCTCCCGTTT TTTACTTTTT CTGTCGATAG

1401 GCGATTGCAG TGGCACTGGC TGGTTTCGCT ACCGTAGCGC AGGCCGACTA
    CGCTAACGTC ACCGTGACCG ACCAAAGCGA TGGCATCGCG TCCGGCTGAT

                                EcoRV
                                ~~~~~
1451 CAAAGATATC GTGATGACCC AGTCTCCAGC AATCATGTCT ACATCTCTAG
    GTTCTATAG CACTACTGGG TCAGAGGTCG TTAGTACAGA TGTAGAGATC

1501 GGGAACGGGT CACCATGACC TGCCTGCCA GTTCAAGTGT AAGTTCCTCT
    CCCTTGCCCA GTGGTACTGG ACGTGACGGT CAAGTTCACA TTCAAGGAGA

1551 TACTTACACT GGTACCAGCA GAAGCCAGGA TCCTCCCCCA AACTCTGGAT
    ATGAATGTGA CCATGGTCGT CTTCGGTCCT AGGAGGGGGT TTGAGACCTA

1601 TTATAGCACA TCCAACCTGG CTTCTGGAGT CCCAACTCGC TTCAGTGGCA

```

Figure 25: pTERMsc2H10myc3sCAM: functional map and sequence (continued)

```

AATATCGTGT AGGTTGGACC GAAGACCTCA GGGTTGAGCG AAGTCACCGT
1651 GTGGGTCTGG GACCTCTTAC TCTCTCACAA TCAGCACCAT GCGGCTGAG
      CACCCAGACC CTGGAGAATG AGAGAGTGTT AGTCGTGGTA CCGCCGACTC
1701 GATGCTGCCA CTTATTACTG CCACCAGTAT CATCGTTTCC CACCCACGTT
      CTACGACGGT GAATAATGAC GGTGGTCATA GTAGCAAAGG GTGGGTGCAA
1751 CGGAGGGGGG ACCAAGCTGG AAATAAAACG GGCTGGTGGT GGTGGTTCTG
      GCCTCCCCC  TGGTTCGACC TTTATTTTGC CCGACCACCA CCACCAAGAC
1801 GCGGCGGCGG CTCCGGTGGT GGTGGTTCTG AAGTTAAACT GGTGAGTCT
      CGCCGCCGCC GAGGCCACCA CCACCAAGAC TTCAATTTGA CCAGCTCAGA
1851 GGAGGAGGCT TGGTGCAACC TGGAGGATCC ATGAAACTCT CCTGTGTTGC
      CCTCCTCCGA ACCACGTTGG ACCTCCTAGG TACTTTGAGA GGACACAACG
1901 CTCTGGAATC ACTTTCAGTA ATTACCGGAT GAACTGGGTC CGCCAGTCTC
      GAGACCTTAG TGAAAGTCAT TAATGGCCTA CTTGACCCAG GCGGTCAGAG
1951 CAGAGAAGGG GCTTGAGTGG GTTGCTGAAA TTAGATTGAA ATCTAATAAT
      GTCTCTTCCC CGAACTCACC CAACGACTTT AATCTAACTT TAGATTATTA
2001 TATGCAACAC ATTATGCGGA GTCTGTGAAA GGGAGGTTCA CCATCTCAAG
      ATACGTTGTG TAATACGCCT CAGACACTTT CCCTCCAAGT GGTAGAGTTC
2051 AGATGATTCC AAAAGTAGTG TCTACCTGCA AATGAACAAC TTAAGAGCTG
      TCTACTAAGG TTTTCATCAC AGATGGACGT TTAGTTGTTG AATTCTCGAC
2101 AAGACACTGG CATTTATTAC TGTAGAGGGG TTTCATATAC TATAGACTAC
      TTCTGTGACC GTAAATAATG ACATCTCCCC AAAGTATATG ATATCTGATG

                                     EcoRI
                                     ~~~~~
2151 TGGGGTCAAG GAACCTCAGT CACAGTCTCC TCAGAATTCC AGCAGAAGCT
      ACCCCAGTTC CTTGGAGTCA GTGTCAGAGG AGTCTTAAGC TCGTCTTCGA
2201 GATCTCTGAG GAAGACCTGT AGGCATGCTT ATTTGTTTGT GAATATCAAG
      CTAGAGACTC CTTCTGGACA TCCGTACGAA TAAACAAACA CTTATAGTTC
2251 GCCAATCGTC TGACCTGCCT CAACCTCCTG TCAATGCTGG CCGCGGCTCT
      CGGTTAGCAG ACTGGACGGA GTTGGAGGAC AGTTACGACC GCCGCCGAGA
2301 GGTGGTGGTT CTGGTGGCGG CTCTGAGGGT GGTGGCTCTG AGGGTGGCGG
      CCACCACCAA GACCACCGCC GAGACTCCCA CCACCGAGAC TCCACCGCC
2351 TTCTGAGGGT GGCGGCTCTG AGGGAGGCGG TTCCGGTGGT GGCTCTGGTT
      AAGACTCCCA CCGCCGAGAC TCCCTCCGCC AAGGCCACCA CCGAGACCAA
2401 CCGGTGATTT TGATTATGAA AAGATGGCAA ACGCTAATAA GGGGGCTATG
      GGCCACTAAA ACTAATACTT TTCTACCGTT TGCGATTATT CCCCCGATAC

```

Figure 25: pTERMsc2H10myc3sCAM: functional map and sequence (continued)

```

2451  ACCGAAAATG CCGATGAAAA CGCGCTACAG TCTGACGCTA AAGGCCAACT
      TGGCTTTTAC GGCTACTTTT GCGCGATGTC AGACTGCGAT TTCCGTTTGA

2501  TGATTCTGTC GCTACTGATT ACGGTGCTGC TATCGATGGT TTCATTGGTG
      ACTAAGACAG CGATGACTAA TGCCACGACG ATAGCTACCA AAGTAACCAC

2551  ACGTTTCCGG CCTTGCTAAT GGTAATGGTG CTACTGGTGA TTTTGCTGGC
      TGCAAAGGCC GGAACGATTA CCATTACCAC GATGACCACT AAAACGACCG

2601  TCTAATTCCC AAATGGCTCA AGTCGGTGAC GGTGATAATT CACCTTTAAT
      AGATTAAGGG TTTACCGAGT TCAGCCACTG CCACTATTAA GTGGAAATTA

2651  GAATAATTTC CGTCAATATT TACCTTCCCT CCCTCAATCG GTTGAATGTC
      CTTATTAAAG GCAGTTATAA ATGGAAGGGA GGGAGTTAGC CAACTTACAG

2701  GCCCTTTTGT CTTTGGCGCT GGTAACCACAT ATGAATTTTC TATTGATTGT
      CGGGAAAACA GAAACCGCGA CCATTTGGTA TACTTAAAAG ATAACATAACA

2751  GACAAAATAA ACTTATTCCG TGGTGTCTTT GCGTTTCTTT TATATGTTGC
      CTGTTTTATT TGAATAAGGC ACCACAGAAA CGCAAAGAAA ATATACAACG

2801  CACCTTTATG TATGTATTTT CTACGTTTGC TAACATACTG CGTAATAAGG
      GTGGAAATAC ATACATAAAA GATGCAAACG ATTGTATGAC GCATTATTCC

      HindIII
      ~~~~~~
2851  AGTCTTGATA AGCTTGACCT GTGAAGTGAA AAATGGCGCA CATTGTGCGA
      TCAGAACTAT TCGAACTGGA CACTTCACTT TTTACCGCGT GTAACACGCT

2901  CATTTTTTTT GTCTGCCGTT TACCGCTACT GCGTCACGGA TCCCCACGCG
      GTAAAAAAA CAGACGGCAA ATGGCGATGA CGCAGTGCCT AGGGGTGCGC

2951  CCCTGTAGCG GCGCATTAAAG CGCGGCGGGT GTGGTGGTTA CGCGCAGCGT
      GGGACATCGC CGCGTAATTC GCGCCGCCCA CACCACCAAT GCGCGTCGCA

3001  GACCGCTACA CTTGCCAGCG CCCTAGCGCC CGCTCCTTTC GCTTTCTTCC
      CTGGCGATGT GAACGGTCGC GGGATCGCGG GCGAGGAAAG CGAAAGAAGG

3051  CTTCTTTTCT CGCCACGTTT GCCGGCTTTC CCCGTCAAGC TCTAAATCGG
      GAAGGAAAGA GCGGTGCAAG CGGCCGAAAG GGGCAGTTCG AGATTTAGCC

3101  GGCATCCCTT TAGGGTTCCG ATTTAGTGCT TTACGGCACC TCGACCCCAA
      CCGTAGGGAA ATCCCAAGGC TAAATCACGA AATGCCGTGG AGCTGGGGTT

3151  AAAACTTGAT TAGGGTGATG GTTCACGTAG TGGGCCATCG CCCTGATAGA
      TTTTGAAC TAATCCACTAC CAAGTGCATC ACCCGGTAGC GGGACTATCT

3201  CGGTTTTTCG CCCTTTGACG TTGGAGTCCA CGTTCTTTAA TAGTGGAAGC
      GCCAAAAGC GGGAAACTGC AACCTCAGGT GCAAGAAATT ATCACCTGAG

```

Figure 25: pTERMsc2H10myc3sCAM: functional map and sequence (continued)

```

3251  TTGTTCCAAA CTGGAACAAC ACTCAACCCT ATCTCGGTCT ATTCTTTTGA
      AACAAGGTTT GACCTTGTTG TGAGTTGGGA TAGAGCCAGA TAAGAAAACCT

3301  TTTATAAGGG ATTTTGCCGA TTTCGGCCTA TTGGTTAAAA AATGAGCTGA
      AAATATTCCC TAAAACGGCT AAAGCCGGAT AACCAATTTT TTA CTGACT

3351  TTTAACAAAA ATTTAACGCG AATTTTAACA AAATATTAAC GTTTACAATT
      AAATTGTTTT TAAATTGCGC TTAA AATTGT TTTATAATTG CAAATGT TAA

3401  TCAGGTGGCA CTTTTCGGGG AAATGTGCGC GGAACCCCTA TTTGTTTATT
      AGTCCACCGT GAAAAGCCCC TTTACACGCG CCTTGGGGAT AAACAAATAA

3451  TTTCTAAATA CATTCAAATA TGTATCCGCT CATGTCGAGA CGTTGGGTGA
      AAAGATTTAT GTAAGTTTAT ACATAGGCGA GTACAGCTCT GCAACCCACT

3501  GGTTCCAACT TTCACCATAA TGAAATAAGA TCACTACCGG GCGTATTTTT
      CCAAGGTTGA AAGTGGTATT ACTTTATTCT AGTGATGGCC CGCATAAAAA

3551  TGAGTTATCG AGATTTTCAG GAGCTAAGGA AGCTAAAATG GAGAAAAAAA
      ACTCAATAGC TCTAAAAGTC CTCGATTCCT TCGATTTTAC CTCTTTTTTT

3601  TCACTGGATA TACCACCGTT GATATATCCC AATGGCATCG TAAAGAACAT
      AGTGACCTAT ATGGTGGCAA CTATATAGGG TTACCGTAGC ATTTCTTGTA

3651  TTTGAGGCAT TTCAGTCAGT TGCTCAATGT ACCTATAACC AGACCGTTCA
      AAAC TCCGTA AAGTCAGTCA ACGAGTTACA TGGATATTGG TCTGGCAAGT

3701  GCTGGATATT ACGGCCTTTT TAAAGACCGT AAAGAAAAAT AAGCACAAGT
      CGACCTATAA TGCCGGAAAA ATTTCTGGCA TTTCTTTTTA TTCGTGTTCA

3751  TTTATCCGGC CTTTATTCAC ATTCTTGCCC GCCTGATGAA TGCTCATCCG
      AAATAGGCCG GAAATAAGTG TAAGAACGGG CGGACTACTT ACGAGTAGGC

3801  GAGTTCCGTA TGGCAATGAA AGACGGTGAG CTGGTGATAT GGGATAGTGT
      CTCAAGGCAT ACCGTTACTT TCTGCCACTC GACCACTATA CCCTATCACA

3851  TCACCCTTGT TACACCGTTT TCCATGAGCA AACTGAAACG TTTTCATCGC
      AGTGGGAACA ATGTGGCAAA AGGTACTCGT TTGACTTTGC AAAAGTAGCG

3901  TCTGGAGTGA ATACCACGAC GATTTCCGGC AGTTTCTACA CATATATTCTG
      AGACCTCACT TATGGTGCTG CTAAAGGCCG TCAAAGATGT GTATATAAGC

3951  CAAGATGTGG CGTGTTACGG TGAAAACCTG GCCTATTTCC CTAAAGGGTT
      GTTCTACACC GCACAATGCC ACTTTTGGAC CGGATAAAGG GATTTCCTAA

4001  TATTGAGAAT ATGTTTTTCG TCTCAGCCAA TCCCTGGGTG AGTTTCACCA
      ATA ACTCTTA TACAAAAAGC AGAGTCGGTT AGGGACCCAC TCAAAGTGGT

4051  GTTTTGATTT AAACGTGGCC AATATGGACA ACTTCTTCGC CCCC GTTTTC
      CAAA ACTAAA TTTGCACCGG TTATACCTGT TGAAGAAGCG GGGGCAAAAG

```

Figure 25: pTERMsc2H10myc3sCAM: functional map and sequence (continued)

|      |            |            |            |             |            |
|------|------------|------------|------------|-------------|------------|
| 4101 | ACCATGGGCA | AATATTATAC | GCAAGGCGAC | AAGGTGCTGA  | TGCCGCTGGC |
|      | TGGTACCCGT | TTATAATATG | CGTTCCGCTG | TTCCACGACT  | ACGGCGACCG |
| 4151 | GATTCAGGTT | CATCATGCCG | TCTGTGATGG | CTTCCATGTC  | GGCAGAATGC |
|      | CTAAGTCCAA | GTAGTACGGC | AGACACTACC | GAAGGTACAG  | CCGTCTTACG |
|      |            | ScaI       |            |             |            |
|      |            | ~~~~~      |            |             |            |
| 4201 | TTAATGAATT | ACAACAGTAC | TGCGATGAGT | GGCAGGGCGG  | GGCGTAATTT |
|      | AATTACTTAA | TGTTGTCATG | ACGCTACTCA | CCGTCCC GCC | CCGCATTAAA |
| 4251 | TTTTAAGGCA | GTTATTGGTG | CCCTTAAACG | CCTGGTGCTA  | CGCCTGAATA |
|      | AAAATTCCGT | CAATAACCAC | GGGAATTTGC | GGACCACGAT  | GCGGACTTAT |
| 4301 | AGTGATAATA | AGCGGATGAA | TGGCAGAAAT | TCGAAAGCAA  | ATTCGACCCG |
|      | TCACTATTAT | TCGCCTACTT | ACCGTCTTTA | AGCTTTCGTT  | TAAGCTGGGC |
| 4351 | GTCGTCGGTT | CAGGGCAGGG | TCGTTAAATA | GCCGCTTATG  | TCTATTGCTG |
|      | CAGCAGCCAA | GTCCCGTCCC | AGCAATTTAT | CGGCGAATAC  | AGATAACGAC |
| 4401 | GTTTACCGGT | TTATTGACTA | CCGGAAGCAG | TGTGACCGTG  | TGCTTCTCAA |
|      | CAAATGGCCA | AATAACTGAT | GGCCTTCGTC | AACTGGCAC   | ACGAAGAGTT |
| 4451 | ATGCCTGAGG | CCAGTTTGCT | CAGGCTCTCC | CCGTGGAGGT  | AATAATTGCT |
|      | TACGGACTCC | GGTCAAACGA | GTCCGAGAGG | GGCACCTCCA  | TTATTAACGA |
| 4501 | CGACATGACC | AAAATCCCTT | AACGTGAGTT | TTCGTTCCAC  | TGAGCGTCAG |
|      | GCTGTACTGG | TTTTAGGGAA | TTGCACTCAA | AAGCAAGGTG  | ACTCGCAGTC |
| 4551 | ACCCCGTAGA | AAAGATCAAA | GGATCTTCTT | GAGATCCTTT  | TTTTCTGCGC |
|      | TGGGGCATCT | TTTCTAGTTT | CCTAGAAGAA | CTCTAGGAAA  | AAAAGACGCG |
| 4601 | GTAATCTGCT | GCTTGCAAAC | AAAAAAACCA | CCGCTACCAG  | CGGTGGTTTG |
|      | CATTAGACGA | CGAACGTTTG | TTTTTTTGGT | GGCGATGGTC  | GCCACCAAAC |
| 4651 | TTTGCCGGAT | CAAGAGCTAC | CAACTCTTTT | TCCGAAGGTA  | ACTGGCTTCA |
|      | AAACGGCCTA | GTTCTCGATG | GTTGAGAAAA | AGGCTTCCAT  | TGACCGAAGT |
| 4701 | GCAGAGCGCA | GATACCAAAT | ACTGTCCTTC | TAGTGTAGCC  | GTAGTTAGGC |
|      | CGTCTCGCGT | CTATGGTTTA | TGACAGGAAG | ATCACATCGG  | CATCAATCCG |
| 4751 | CACCACTTCA | AGAACTCTGT | AGCACCGCCT | ACATACCTCG  | CTCTGCTAAT |
|      | GTGGTGAAGT | TCTTGAGACA | TCGTGGCGGA | TGTATGGAGC  | GAGACGATTA |
| 4801 | CCTGTTACCA | GTGGCTGCTG | CCAGTGGCGA | TAAGTCGTGT  | CTTACCGGGT |
|      | GGACAATGGT | CACCGACGAC | GGTCACCGCT | ATTACGCACA  | GAATGGCCCA |
| 4851 | TGGACTCAAG | ACGATAGTTA | CCGGATAAGG | CGCAGCGGTC  | GGGCTGAACG |
|      | ACCTGAGTTC | TGCTATCAAT | GGCCTATTCC | GCGTCGCCAG  | CCCGACTTGC |
| 4901 | GGGGGTTCGT | GCACACAGCC | CAGCTTGGAG | CGAACGACCT  | ACACCGAACT |

Figure 25: pTERMsc2H10myc3sCAM: functional map and sequence (continued)

```
          CCCCCAAGCA CGTGTGTCGG GTCGAACCTC GCTTGCTGGA TGTGGCTTGA
4951  GAGATACCTA CAGCGTGAGC TATGAGAAAG CGCCACGCTT CCCGAAGGGA
      CTCTATGGAT GTCGCACTCG ATACTCTTTC GCGGTGCGAA GGGCTTCCCT
5001  GAAAGGCGGA CAGGTATCCG GTAAGCGGCA GGGTCGGAAC AGGAGAGCGC
      CTTTCCGCCT GTCCATAGGC CATTGCGCCG CCCAGCCTTG TCCTCTCGCG
5051  ACGAGGGAGC TTCCAGGGGG AAACGCCTGG TATCTTTATA GTCCTGTCGG
      TGCTCCCTCG AAGGTCCCCC TTTGCGGACC ATAGAAATAT CAGGACAGCC
5101  GTTTCGCCAC CTCTGACTTG AGCGTCGATT TTTGTGATGC TCGTCAGGGG
      CAAAGCGGTG GAGACTGAAC TCGCAGCTAA AAACACTACG AGCAGTCCCC
5151  GGCGGAGCCT ATGGAAAAAC GCCAGCAACG CGGCCTTTTT ACGGTTCCCTG
      CCGCCTCGGA TACCTTTTTC CGGTCGTTGC GCCGGAAAAA TGCCAAGGAC
5201  GCCTTTTGCT GGCCTTTTGC TCACATG
      CGGAAAACGA CCGGAAAACG AGTGTAC
```

Table 1: Phagemids Constructed for Experiments 2 and 3

| Name         | FLAG | His6 | gIII | Size<br>(bp) | Insert    | REN1  | REN2    | Resistance |
|--------------|------|------|------|--------------|-----------|-------|---------|------------|
| pING1-A1     | -    | +    | -    | 3783         | His       | EcoRV | SmaI    | Ap         |
| pING1-A2     | -    | -    | -    | 3795         | Strep-tag | EcoRV | SmaI    | Ap         |
| pING3-A1     | +    | +    | -    | 3792         | His       | EcoRV | SmaI    | Ap         |
| pING3-A2     | +    | -    | -    | 3804         | Strep-tag | EcoRV | SmaI    | Ap         |
| pONG3-A      | +    | -    | +    | 4278         | -         | EcoRV | SmaI    | Ap         |
| pYANG3-A     | +    | -    | +    | 4404         | Jun       | EcoRV | EcoRI   | Ap         |
| pYANG3-Ape2  | +    | -    | +    |              | pep2      | XbaI  | HindIII | Ap         |
| pYANG3-Ape3  | +    | -    | +    |              | pep3      | XbaI  | HindIII | Ap         |
| pYANG3-Ape10 | +    | -    | +    |              | pep10     | XbaI  | HindIII | Ap         |
| pING1-C1     | -    | +    | -    | 3853         | His       | EcoRV | SmaI    | Cm         |
| pING1-C2     | -    | -    | -    | 3865         | Strep-tag | EcoRV | SmaI    | Cm         |
| pING3-C1     | +    | +    | -    | 3862         | His       | EcoRV | SmaI    | Cm         |
| pING3-C2     | +    | -    | -    | 3874         | Strep-tag | EcoRV | SmaI    | Cm         |
| pYING3-C1    | +    | +    | -    | 3994         | Fos       | EcoRV | EcoRI   | Cm         |
| pYING3-C2    | +    | +    | -    | 4315         | p75       | EcoRV | EcoRI   | Cm         |
| pYING3-C3    | +    | +    | -    | 4240         | IL-16     | EcoRV | EcoRI   | Cm         |

Table 2: Results of Experiment 2 (see Figure 7)

Table 2a: Combination of phagemids present in initial library ( $\alpha$ )

|    | Combination             | Clone(s) |
|----|-------------------------|----------|
| 1. | pYING1-C2 + pYANG3-ApeX | 6        |
| 2. | pYING1-C1 + pYANG3-A    | 1        |
| 3. | pYING1-C1 + pYANG3-ApeX | 1        |
| 4. | pYING1-C2 + pYANG3-A    | 1        |
| 5. | pYING1-C2 + ?           | 1        |

Table 2b: Combination of phagemids present after selection ( $\beta$ )

|    | Combination             | Clone(s) |
|----|-------------------------|----------|
| 1. | pYING1-C2 + pYANG3-ApeX | 1        |
| 2. | pYING1-C1 + pYANG3-A    | 9        |



Table 3: Results of Experiment 4 (see Figure 19)

Table 3a: Identification of phage/plasmid present in individual clones

| Combination              | Clone(s)                         |
|--------------------------|----------------------------------|
| fhag1A + pUC19/IMPhag    | #9                               |
| fpep3_1b + pUC18/IMP-p75 | #1,#3,#5,#6,#7,#13,#15,#19       |
| fpep3_1b + pUC19/IMPhag  | #14                              |
| unusual DNA              | #2,#4,#8,#10,#11,#12,#16,#17,#18 |

Table 3b: Test for infectivity of individual clones

| Clone # | Titer (transducing units/ml) |
|---------|------------------------------|
| 1       | $2 \times 10^4$              |
| 2       | 31                           |
| 3       | $1 \times 10^5$              |
| 4       | $1 \times 10^5$              |
| 5       | $1 \times 10^5$              |
| 6       | $2 \times 10^3$              |
| 7       | $1 \times 10^4$              |
| 8       | $1 \times 10^5$              |
| 9       | $1 \times 10^6$              |
| 10      | $1 \times 10^4$              |
| 11      | $1 \times 10^3$              |
| 12      | $1 \times 10^4$              |
| 13      | $3 \times 10^3$              |
| 14      | < 10                         |
| 15      | $5 \times 10^4$              |
| 16      | $1 \times 10^4$              |
| 17      | $5 \times 10^2$              |
| 18      | $1 \times 10^4$              |
| 19      | $1 \times 10^5$              |

# INTERNATIONAL SEARCH REPORT

Intern al Application No  
PCT/EP 97/00931

## A. CLASSIFICATION OF SUBJECT MATTER

IPC 6 C12N15/10 C12N15/12 C12N15/24 C12N15/62 C12N15/70

According to International Patent Classification (IPC) or to both national classification and IPC

## B. FIELDS SEARCHED

Minimum documentation searched (classification system followed by classification symbols)

IPC 6 C12N

Documentation searched other than minimum documentation to the extent that such documents are included in the fields searched

Electronic data base consulted during the international search (name of data base and, where practical, search terms used)

## C. DOCUMENTS CONSIDERED TO BE RELEVANT

| Category * | Citation of document, with indication, where appropriate, of the relevant passages  | Relevant to claim No.                |
|------------|---|--------------------------------------|
| X          | WO 93 19172 A (CAMBRIDGE ANTIBODY TECH<br>;MEDICAL RES COUNCIL (GB)) 30 September<br>1993<br><br>see the whole document<br>---    | 1-7,11,<br>12,<br>19-27,<br>29,30,32 |
| A          | EP 0 614 989 A (MORPHOSYS GES FUER<br>PROTEINOPTI) 14 September 1994<br>cited in the application<br>see the whole document<br>--- | 1-34                                 |
| A          | DE 41 22 598 C (DKFZ) 30 July 1992<br>see the whole document<br>---   | 1-34                                 |
|            | -/-   |                                      |

☒ Further documents are listed in the continuation of box C.

☒ Patent family members are listed in annex.

### \* Special categories of cited documents :

- "A" document defining the general state of the art which is not considered to be of particular relevance
- "E" earlier document but published on or after the international filing date
- "L" document which may throw doubts on priority claim(s) or which is cited to establish the publication date of another citation or other special reason (as specified)
- "O" document referring to an oral disclosure, use, exhibition or other means
- "P" document published prior to the international filing date but later than the priority date claimed

- "T" later document published after the international filing date or priority date and not in conflict with the application but cited to understand the principle or theory underlying the invention
- "X" document of particular relevance; the claimed invention cannot be considered novel or cannot be considered to involve an inventive step when the document is taken alone
- "Y" document of particular relevance; the claimed invention cannot be considered to involve an inventive step when the document is combined with one or more other such documents, such combination being obvious to a person skilled in the art
- "A" document member of the same patent family

Date of the actual completion of the international search

19 June 1997

Date of mailing of the international search report

03.07.97

Name and mailing address of the ISA

European Patent Office, P.B. 5818 Patentlaan 2  
NL - 2280 HV Rijswijk  
Tel. (+ 31-70) 340-2040, Tx. 31 651 epo nl,  
Fax: (+ 31-70) 340-3016

Authorized officer

HORNIG, H

## INTERNATIONAL SEARCH REPORT

Intern. Appl. No.

PCT/EP 97/00931

## C.(Continuation) DOCUMENTS CONSIDERED TO BE RELEVANT

| Category | Citation of document, with indication, where appropriate, of the relevant passages  | Relevant to claim No. |
|----------|---|-----------------------|
| A        | <p>GENE,<br/>vol. 137, 1993, ELSEVIER SCIENCE<br/>PUBLISHERS, B.V., AMSTERDAM, NL;<br/>pages 69-75, XP002033343<br/>R. CAMERI AND M. SUTER: "Display of<br/>biological active proteins on the surface<br/>of filamentous phages: a cDNA cloning<br/>system for selection of functional gene<br/>products linked to genetic information<br/>responsible for their production"<br/>cited in the application<br/>see the whole document</p> <p>---</p> | 1-34                  |
| A        | <p>NUCLEIC ACIDS RESEARCH,<br/>vol. 22, no. 25, 1994, IRL PRESS<br/>LIMITED, OXFORD, ENGLAND,<br/>pages 5761-5762, XP002033344<br/>K. GRAMATIKOFF ET AL.: "Direct<br/>interaction rescue, a novel filamentous<br/>phage technique to study protein - protein<br/>interactions"<br/>cited in the application<br/>see the whole document</p> <p>---</p>   | 1-34                  |
| A        | <p>NATURE BIOTECHNOLOGY,<br/>vol. 13, April 1995, NATURE PUBL. CO., NEW<br/>YORK, US,<br/>pages 373-377, XP002033345<br/>D. NERI ET AL.: "Calmodulin as a<br/>versatile tag for antibody fragments"<br/>see the whole document</p> <p>---</p>   | 1-34                  |
| A        | <p>NATURE BIOTECHNOLOGY,<br/>vol. 13, April 1995, NATURE PUBL. CO., NEW<br/>YORK, US,<br/>pages 366-372, XP002033346<br/>Z. LU ET AL.: "Expression of thioredoxin<br/>random peptide libraries on the<br/>Escherichia coli cell surface as<br/>functional fusions to flagellin: A system<br/>designed for exploring protein-protein<br/>interactions"<br/>cited in the application<br/>see the whole document</p> <p>---</p>                        | 1-34                  |
| A        | <p>FEBS LETTERS,<br/>vol. 377, no. 2, 18 December 1995,<br/>ELSEVIER, AMSTERDAM, NL,<br/>pages 227-231, XP002033347<br/>C. KREBBER ET AL.: "Co-selection of<br/>cognate antibody-antigen pairs by<br/>selectively-infective phage"<br/>cited in the application<br/>see the whole document</p> <p>---</p> <p style="text-align: center;">-/--</p>   | 1-34                  |

# INTERNATIONAL SEARCH REPORT

Int'l. Application No  
PCT/EP 97/00931

## C.(Continuation) DOCUMENTS CONSIDERED TO BE RELEVANT

| Category | Citation of document, with indication, where appropriate, of the relevant passages   | Relevant to claim No. |
|----------|--|-----------------------|
| E        | <p>WO 97 10330 A (BIOINVENT INT AB ;KIDDLE<br/>SIMON JOHN (GB); BORREBAECK CARL (SE); MA)<br/>20 March 1997<br/>see the whole document<br/>-----</p> | 1,2                   |

Form PCT/ISA/210 (continuation of second sheet) (July 1992)

# INTERNATIONAL SEARCH REPORT

Information on patent family members

Intern. Application No

PCT/EP 97/00931

| Patent document<br>cited in search report | Publication<br>date | Patent family<br>member(s) | Publication<br>date |
|---|---------------------|----------------------------|---------------------|
| WO 9319172 A                              | 30-09-93            | AU 665190 B                | 21-12-95            |
|   |                     | AU 1693892 A               | 30-12-92            |
|   |                     | AU 665025 B                | 14-12-95            |
|   |                     | AU 2593392 A               | 27-04-93            |
|   |                     | AU 665221 B                | 21-12-95            |
|   |                     | AU 3089092 A               | 28-06-93            |
|   |                     | AU 673515 B                | 14-11-96            |
|   |                     | AU 3763893 A               | 21-10-93            |
|   |                     | CA 2109602 A               | 26-11-92            |
|   |                     | CA 2119930 A               | 01-04-93            |
|   |                     | CA 2124460 A               | 10-06-93            |
|   |                     | CA 2131151 A               | 30-09-94            |
|   |                     | EP 0585287 A               | 09-03-94            |
|   |                     | EP 0605522 A               | 13-07-94            |
|   |                     | EP 0616640 A               | 28-09-94            |
|   |                     | EP 0656941 A               | 14-06-95            |
|   |                     | WO 9220791 A               | 26-11-92            |
|   |                     | WO 9306213 A               | 01-04-93            |
|   |                     | WO 9311236 A               | 10-06-93            |
|   |                     | JP 6510671 T               | 01-12-94            |
|   |                     | JP 6508511 T               | 29-09-94            |
|   |                     | JP 7502167 T               | 09-03-95            |
|   |                     | JP 7505055 T               | 08-06-95            |
|   |                     | US 5565332 A               | 15-10-96            |
| EP 0614989 A                              | 14-09-94            | CA 2115811 A               | 18-08-94            |
|   |                     | DE 614989 T                | 28-09-95            |
|   |                     | JP 7039381 A               | 10-02-95            |
|   |                     | US 5514548 A               | 07-05-96            |
| DE 4122598 C                              | 30-07-92            | WO 9301287 A               | 21-01-93            |
|   |                     | EP 0547200 A               | 23-06-93            |
|   |                     | JP 6501395 T               | 17-02-94            |
|   |                     | US 5591604 A               | 07-01-97            |
| WO 9710330 A                              | 20-03-97            | AU 6939196 A               | 01-04-97            |